

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
14 March 2002 (14.03.2002)

PCT

(10) International Publication Number
WO 02/20814 A1

(51) International Patent Classification⁷: **C12N 15/86**,
15/861, 5/10, 15/11, 15/63, 15/64, 15/65, A61K 48/00

Eun-Chung [KR/US]; 91 Waltham Street #1, Boston, MA 02118 (US). WANG, Xinzong [CN/US]; 4 Woodmere Road, Framingham, MA 01701 (US).

(21) International Application Number: PCT/US01/27682

(74) Agent: ELBING, Karen, L.; Clark & Elbing LLP, 176 Federal Street, Boston, MA 02110-2214 (US).

(22) International Filing Date:
7 September 2001 (07.09.2001)

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(25) Filing Language: English

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(26) Publication Language: English

Published:

(30) Priority Data:
60/231,053 8 September 2000 (08.09.2000) US
60/246,904 8 November 2000 (08.11.2000) US

- with international search report
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments
- with sequence listing part of description published separately in electronic form and available upon request from the International Bureau

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US 60/231,053 (CON)
Filed on 8 September 2000 (08.09.2000)
US 60/246,904 (CON)
Filed on 8 November 2000 (08.11.2000)

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(71) Applicant (*for all designated States except US*): **THE GENERAL HOSPITAL CORPORATION [US/US]**; 55 Fruit Street, Boston, MA 02214 (US).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **SEED, Brian** [US/US]; 9 Hawthorne Place #5J, Boston, MA 02114 (US). **WRIGHT FREEMAN, Mason** [US/US]; 203 Lincoln Road, Lincoln, MA 01773 (US). **KOVTON, Alexander** [US/US]; 156 Arlington Street, Acton, MA 01720 (US). **MURAKAWA, Masahiro** [JP/US]; 506 Beacon Street, Apt. 6, Boston, MA 02115 (US). **PARK,**

WO 02/20814 A1

(54) Title: SELF-REARRANGING DNA VECTORS

(57) Abstract: Disclosed are replicatable viral DNA vectors encoding a site-specific DNA-altering enzyme and a DNA target recognized by the enzyme, the enzyme selectively converting, in a cell expressing the enzyme, the DNA vector to a rearranged form. The invention further relates to methods for assembling recombinant adenoviral DNAs. These methods include the steps of: providing a first linearized DNA vector including a restriction site and a cos site and a second linearized DNA vector including the restriction site, an adenoviral nucleic molecule, and a cos site; and ligating the first and second linearized DNA vectors, the ligation assembling a recombinant adenoviral DNA.

SELF-REARRANGING DNA VECTORS

Background of the Invention

The invention relates to DNA vectors.

10 Mammalian cell expression vectors based on DNA viruses have been widely discussed as gene delivery vehicles for genetic therapy. Among the different DNA viruses proposed for this purpose have been adenoviruses, baculovirus, Epstein Barr virus, and herpes simplex virus. In addition other smaller viruses that have an intranuclear phase in which the viral genome is present as a double stranded DNA, such 15 as retroviruses and parvoviruses, have been proposed as gene delivery vehicles.

Adenoviral vectors (AdV), for example, have a recognized potential for gene delivery, founded in their broad host range, robust growth in culture, and capacity to infect mitotically quiescent cells (Graham and Prevec, Manipulation of adenovirus vectors, p. 109-128, *In E. J. Murray (ed.), Methods in Molecular Biology*, vol. 7, 20 Humana, Clifton, NJ, 1991; Trapnell and Gorziglia, *Curr. Opin. Biotechnol.* 5:617-625, 1994). AdV can be propagated in a helper cell line, 293, a human embryonic kidney cell line transformed by adenovirus type 5 (Graham et al., *J. Gen. Virol.* 36:59-72, 1994). 293 cells express the viral E1 gene products (E1a and E1b) that are the master regulatory proteins for subsequent viral gene expression. E1 deleted viruses can propagate in 293 25 cells, but not in other cells. Although it would be expected that E1 deleted viruses lack the machinery to express viral genes, several studies have demonstrated that cellular E1-like components can stimulate viral gene expression (Imperiale et al., *Mol. Cell. Biol.* 4:867-74, 1984; Oncleaq et al., *J. Virol.* 62:4533-7, 1988; Spergel et al., *J. Virol.* 66:1021-30, 1992). The expression of these viral genes results in the relatively rapid 30 elimination of transduced cells *in vivo* as a result of cytotoxic T cell responses (Yang et al., *Immunity* 1:433-42, 1994; Yang et al., *Gene Ther.* 3:137-44, 1996; Yang et al., *J. Virol.* 69:2004-15, 1995).

Thus attention has been focused on eliminating the remaining vestiges of viral expression. Viral genes that have been deleted for this purpose include the gene for E4 35 proteins (Armentano et al., *Hum. Gene Ther.* 6:1343-53, 1995; Kochanek et al., *Proc.*

Natl. Acad. Sci. USA 93:5731-6, 1996; and Yeh et al., J. Virol. 70:559-565, 1996), DNA binding protein (Engelhardt et al., Proc. Natl. Acad. Sci. USA 21:6196-6200, 1994; and Gorziglia et al., J. Virol. 70:4173-8, 1996), DNA polymerase (Amalfitano et al., J. Virol. 72:926-33, 1998), and the preterminal protein (Schaack et al., Proc. Natl. Acad. Sci. USA 93:14686-91, 1996). The most aggressive approach has been the creation of helper virus-dependent vectors that lack all viral genes (Hardy et al., J. Virol. 71:1842-9, 1997; Kochanek et al., Proc. Natl. Acad. Sci. USA 93:5731-6, 1996; Lieber et al., J. Virol. 70:8944-60, 1996; Mitani et al., Proc. Natl. Acad. Sci. USA 92:3854-8, 1995; and Parks et al., Proc. Natl. Acad. Sci. USA 93:13565-13570, 1996). These vectors have high capacity, evoke reduced cellular immune responses and show prolonged expression *in vivo* (Morsy et al., Proc. Natl. Acad. Sci. USA 95:7866-71, 1998). However to deploy these viruses on the scale required for human clinical application presents major challenges because a cesium chloride (CsCl) gradient is needed to remove the helper virus.

15

Summary of the Invention

In one aspect, the invention features a replicatable viral DNA vector encoding a site-specific DNA-altering enzyme and a DNA target recognized by said enzyme, said enzyme selectively converting, in a cell expressing said enzyme, said DNA vector to a rearranged form.

In preferred embodiments, the rearranged form includes an autonomously replicating episome and a linear DNA product. In other preferred embodiments, the vector comprises adenoviral DNA.

In yet other preferred embodiments, the vector includes a genetically-engineered recombination site (such as a target of Cre or FLP). Preferably, such a recombination site includes a recognition sequence of a site-specific DNA altering enzyme.

In another preferred embodiment, the site-specific DNA altering enzyme is a recombinase (such as Cre or FLP) or an integrase. Preferably, such an enzyme is functional in a mammalian cell. Preferred embodiments of the vector also include an origin of replication that functions in a mammalian cell (such as an Epstein Barr Virus replicon). Moreover, the vector typically includes a gene of interest (such as a therapeutic gene that encodes a protein or polypeptide or an RNA product).

In another aspect, the invention features a method for assembling a recombinant adenoviral DNA. The method, in general, includes the steps of: (a) providing a first linearized DNA vector comprising a restriction site and a cos site and a second linearized DNA vector comprising the restriction site, an adenoviral nucleic acid molecule, and a cos site; and (b) ligating the first and second linearized DNA vectors, the ligation assembling a recombinant adenoviral DNA.

5 In preferred embodiments, the first linearized DNA vector comprises a selectable marker (such as a gene encoding a polypeptide that confers, on a host cell expressing such a polypeptide, resistance to an antibiotic). In other preferred embodiments, the first linearized DNA vector includes an adenoviral left-end inverted terminal repeat; a gene of interest; or both. In still other preferred embodiments, the second linearized DNA vector includes a selectable marker. Preferably, the second linearized DNA vector includes an adenoviral right-end inverted terminal repeat.

10 The method further includes packaging the assembled adenoviral DNA into a phage and infecting a host cell. Typically the first and second linearized DNAs include cosmid vector DNA. In addition, such adenoviral DNA is typically flanked by cleavage sites (such as intron endonuclease cleavage sites).

15 In another aspect, the invention features an adenovirus producer cell having a nucleic acid molecule that expresses a dominant negative site-specific DNA-altering enzyme. In preferred embodiments, the site-specific DNA altering enzyme is a dominant negative recombinase (for example, a Cre recombinase such as CreY324C or a Flp recombinase). Exemplary adenovirus producer cells include, without limitation, 293 human embryonic kidney cells, per.C6 cells, and N52 cells.

20 In yet another aspect, the invention features a vector comprising, in the 5' to 3' direction, a first genetically engineered *cis*-acting target recognized by a site-specific DNA altering enzyme; a gene of interest; a lineage-specific gene promoter; a second genetically engineered *cis*- acting target recognized by a site-specific DNA altering enzyme; and a nucleic acid molecule encoding a site-specific DNA altering enzyme.

25 In still another aspect, the invention features a vector including, in the 5' to 3' direction, a first genetically engineered *cis*-acting target recognized by a site-specific DNA altering enzyme; a gene of interest; a bi-directional promoter, comprising a second genetically engineered *cis*-acting target recognized by a site-specific DNA altering enzyme; and a nucleic acid molecule encoding a site-specific DNA altering enzyme.

In related aspects, the invention features a method of gene therapy including the administration to a patient in need of gene therapy a therapeutically effective amount of the vector of the invention, which is expressed in the patient. The invention further relates to a population of cells transfected with the vector of the invention.

5 Accordingly, the invention further relates to the use of a recombinant viral vector or use of a recombinant viral particle for gene therapy. Such vectors and viral particles may be introduced either *in vitro* into a host cell removed from the patient, or directly *in vivo*, into the body to be treated, according to standard methods known in the art.

The invention also relates to a pharmaceutical composition that includes a
10 therapeutically effective amount of a recombinant viral vector or viral particle prepared according to the methods disclosed herein, in combination with a vehicle that is acceptable from a pharmaceutical standpoint. Such a pharmaceutical composition may be prepared according to the techniques commonly employed and administered by any known administration route, for example systemically (in particular, by intravenous,
15 intratracheal, intraperitoneal, intramuscular, subcutaneous, intratumoral, or intracranial routes) or by aerosolization or intrapulmonary administration.

One skilled in the art will appreciate that suitable methods of administering a vector (particularly an adenoviral vector) of the present invention to an animal for purposes of gene therapy, chemotherapy, and vaccination are available, and, although
20 more than one route can be used for administration, one particular route may provide a more immediate and more effective reaction than another. Pharmaceutically acceptable excipients also are well known to those who are skilled in the art, and are readily available. The choice of excipient will be determined, in part, by the particular method used to administer the recombinant vector or particle. Accordingly, there are a wide
25 variety of suitable formulations for use in the context of the present invention.

By "recombinant DNA vector" is meant a DNA sequence containing a desired sequence (such as a gene of interest) and an appropriate regulatory element(s) necessary for the expression of the operably linked sequence in a particular host organism (such as a mammal).

30 By "operably linked" is meant that a gene and a regulatory element(s) are connected to permit gene expression when the appropriate molecules (for example, transcriptional activator proteins) are bound to the regulatory sequence(s).

By "regulatory element" is meant a genetic element that controls some aspect of the expression of a nucleic acid sequence. For example, a promoter is a regulatory element that facilitates the initiation of transcription of an operably linked coding region. Other genetic regulatory elements include, without limitation, splicing signals, 5 polyadenylation signals, and termination signals. For example, transcriptional regulatory elements in eukaryotes include promoter and enhancer elements. Promoters and enhancers include arrays of DNA sequences that interact directly or indirectly with cellular proteins involved in transcription. Promoter and enhancer elements have been isolated from a variety of eukaryotic sources including genes in mammalian cells and 10 viruses.

By "transfection" is meant the introduction of foreign DNA into eukaryotic cells. Transfection is typically accomplished by a variety of means known in the art including, without limitation, calcium phosphate-DNA co-precipitation, DEAE-dextran-mediated transfection, electroporation, microinjection, liposome fusion, lipofection, protoplast 15 fusion, and biolistics.

By "stably transfected" is meant the introduction of foreign DNA into the genome of the transfected cell. In general, transfer and expression of transgenes in mammalian cells are now routine practices to those skilled in the art, and have become major tools to carry out gene expression studies and to generate vectors useful in gene 20 therapy.

By "gene of interest" is meant a gene inserted into a vector whose expression is desired in a host cell. Genes of interest include, without limitation, genes having therapeutic value, as well as reporter genes. A variety of such genes are useful in the invention, including genes of interest encoding a protein, which provides a therapeutic 25 function. In addition, the gene of interest, if a therapeutic gene, can render its effect at the level of RNA, for instance, by encoding an antisense message or ribozyme, a protein which affects splicing or 3' processing (e.g., polyadenylation), or it can encode a protein which acts by affecting the level of expression of another gene within the cell (i.e., where gene expression is broadly considered to include all steps from initiation of 30 transcription through production of a processed protein), for example, by mediating an altered rate of mRNA accumulation, an alteration of mRNA transport, and/or a change in post-transcriptional regulation.

By "reporter gene" is meant a gene sequence that encodes a reporter molecule (including an enzyme). A "reporter molecule" is detectable in any detection system, including, but not limited to, enzyme (e.g., ELISA, as well as enzyme-based histochemical assays), fluorescent, radioactive, and luminescent systems. Exemplary reporter gene systems include the *E. coli* beta-galactosidase or glucuronidase genes, green fluorescent protein (GFP), blue fluorescent protein (BFP), the human placental alkaline phosphatase gene, the chloramphenicol acetyltransferase (CAT) gene; other reporter genes are known in the art and may be employed as desired.

By "transgene" is meant any piece of DNA, which is inserted by artifice into a cell, and becomes part of the genome of the organism, which develops from that cell. Such a transgene may include a gene that is partly or entirely heterologous (i.e., foreign) to the transgenic organism, or may represent a gene homologous to an endogenous gene of the organism.

By "transgenic" is meant any cell that includes a DNA sequence, which is inserted by artifice into a cell and becomes part of the genome of the organism, which develops from that cell.

By "polypeptide" is meant any chain of amino acids, regardless of length or post-translational modification (for example, glycosylation or phosphorylation).

By "derived from" is meant isolated from or having the sequence of a naturally occurring sequence (e.g., a cDNA, genomic DNA, synthetic, or combination thereof).

By "nucleic acid" is meant a polynucleotide (DNA or RNA).

By "gene" is meant any nucleic acid sequence coding for a protein or an RNA molecule.

By "gene product" is meant either an untranslated RNA molecule transcribed from a given gene or coding sequence (such as, mRNA or antisense RNA) or the polypeptide chain translated from the mRNA molecule transcribed from the given gene or coding sequence. Nucleic acids according to the invention can be wholly or partially synthetically made, can comprise genomic or complementary DNA (cDNA) sequences, or can be provided in the form of either DNA or RNA.

The presently claimed invention affords a number of advantages. For example, applicants' gene therapy vehicles particularly those based on recombinant adenoviruses, minimize the propensity of the vectors to activate host immune surveillance, and thereby maximize the persistence for the DNA transduced. The invention therefore facilitates

the development of gene delivery vectors designed to enhance persistence of virally delivered genes and evade the cellular immune response by severing the connection between the sole adenoviral enhancer and the sequences encoding potentially antigenic viral proteins.

5 As described in more detail below, the mechanism by which this is accomplished differs significantly from any other previous approaches. For example, to reduce the immunogenicity of vectors it is widely acknowledged that some intervention, such as the removal of key genes, or the prevention of their expression in the cells targeted for therapy, is important; however, many related approaches are directed at the host and
10 have generally focused on the selective induction of tolerance to adenoviral antigens, or similar strategies directed at inducing a temporally restricted or antigen-specific compromise of the immune system.

15 In addition, the poor persistence of transduced DNA appears to be due in part to immunological rejection of transduced cells and to the inability of the viral DNA to replicate, a feature generally inherent in the design of adenoviral vectors, but one which is not associated with applicants' claimed gene therapy vehicles.

20 Moreover, some contemporary adenoviral vectors are designed to propagate in specific host cells which provide essential replication factors *in trans*. These vectors are typically based on cell lines which express the master regulatory proteins of the E1 complex, which are required for induction of adenoviral DNA replication. In cells expressing E1 genes, the best studied of which is a human embryonic kidney cell line transformed by DNA from human adenovirus 5 (called HEK293, or simply 293), viruses lacking E1 genes propagate well. Such viruses do not propagate on cell lines which do not express E1, and do not generally propagate well in the target cells to which the
25 therapeutic gene is to be delivered. Cells transduced with E1-deleted adenovirus vectors also do not express high levels of viral genes in the absence of E1. However, the weak residual expression that remains in such vectors appears to be sufficient to induce cellular immune responses that contribute to the destruction of the transduced cells.

30 In addition, the gene therapy vectors claimed herein are hybrid vectors capable of self-rearrangement to form circular and linear DNA products. The linear DNA has a compromised ability to express adenoviral genes, and therefore has a lower immunological profile. And the circular DNA behaves like a mammalian plasmid, encoding the gene of interest and persisting by autonomous replication in the nucleus.

For example, the circularization of an adenoviral vector via the action of Cre recombinase beneficially places a gene of interest (for example, a therapeutic gene) on a self-replicating episome. Vector circularization occurs in a tissue-targeted manner, for example, as a result of the activation of a synthetic liver-specific promoter upstream of 5 the recombinase Cre. Once circularized, the EBV replicon in the episome confers improved persistence on the therapeutic gene as detected by reporter gene expression and direct assay for the presence of vector DNA sequences.

Furthermore, the invention eliminates the requirement for a helper virus, thus avoiding two potential limitations of that system. First, the continuous expression of Cre 10 recombinase may lead to toxicity in host cells, either as a direct consequence of the protein's activity or via its immunogenicity. Second, the Cre helper virus may itself produce antigenic viral proteins that contribute to the immunologic elimination of infected host cells. In contrast, the self-resolving adenovirus/EBV vector system disclosed herein advantageously provides no alternative source of viral proteins, and Cre 15 expression is terminated upon rearrangement.

In addition, the invention described herein provides tools for analyzing the roles of the enhancer in viral gene regulation and virus growth.

The invention also provides a convenient general system for creating recombinant adenoviruses, which increase their attractiveness as gene transduction tools 20 for basic research. The system, for example, employs two conventional plasmid vectors and a λ phage packaging step. The entire recombinant AdV genome is assembled into a single cosmid that is easily amplified in *E. coli*. The use of intron endonuclease recognition sequences flanking the ITRs enhances virus production while simplifying insertion of therapeutic gene sequences into the pLEP shuttle plasmid. The convenience 25 of this vector system has facilitated the construction of over two hundred recombinant viruses to date.

Other embodiments and advantages of the invention will be apparent from the detailed description thereof, and from the claims.

30

Brief Description of the Drawings

FIGURE 1A is a schematic diagram of the structure of an adenoviral type A vector and its fate in a target cell. enh refers to the Ad2 enhancer; GFP refers to the marker gene green fluorescent protein; EBV refers to the Epstein Barr Virus replicon;

TetO₇ refers to a heptamer of Tet operator; TetR refers to the Tet repressor; VP16 refers to the viral protein 16 of Herpes simplex virus, SD refers to the splice donor site; and SA refers to the splice acceptor site.

FIGURE 1B is a schematic diagram of the structure of an adenoviral type B
5 vector and its fate in a target cell. enh refers to the Ad2 enhancer; GFP refers to the
marker gene green fluorescent protein; EBV refers to the Epstein Barr Virus replicon;
SD refers to the splice donor site; and SA refers to the splice acceptor site.

FIGURE 2A shows a schematic diagram of the pLEP cosmid polylinker region
and its position relative to the adenoviral left ITR. The adenovirus enhancer/packaging
10 sequence (ψ) is boxed.

FIGURE 2B is a schematic diagram showing the generation of a single cosmid
encoding the AdV genome by the direct ligation of two smaller plasmids. A gene
expression unit, CMVGFP, was inserted into the pLEP cosmid at the polylinker region.
pLEP and pREP cosmids were digested with an intron endonuclease (PI-PspI), ligated,
15 and packaged *in vitro* to generate pAd2CMVGFP. This DNA was then digested with
another intron endonuclease (I-CeuI) to expose the ITRs at both ends of the viral
genome. Finally, cosmid digestion mixtures were transfected into 293 cells. Plaques
generated by recombinant viruses are detected in 7-10 days.

FIGURE 3A shows the restriction analysis of cosmids carrying the full length
20 AdV DNA showing uniform generation of the desired vector DNA. 2 μ g DNA samples
from four pAd2-7CMVGFP colonies were digested with Bgl II, resolved on a 1%
agarose gel and stained with ethidium bromide. The predicted sizes of the DNA
fragments are: 13261, 7684, 5228, 5088, 2284, 1757, 1549, 1270, 351, and 275 base
pairs (bp). The 5228 and 5088 fragments appear as a doublet, and the 351 and 275 bp
25 fragments are too small to be seen on the gel.

FIGURE 3B shows the release of the recombinant Ad DNA from cosmids by I-
CeuI digestion. 2 μ g of pAd2-7CMV DNA from two clones was digested with I-CeuI.
Arrows indicated the position of the released recombinant AdV DNA and the vector
fragments of approximately 35 kb and 5 kb, respectively.

30 FIGURE 4A shows the appearance of plaques in 293 cells transfected with 10 μ g
of pIAdGFPB with no ITR exposed (undigested), one ITR exposed (BsaBI or I-CeuI), or
both ITRs exposed (BsaBI plus I-CeuI). Values represent the mean plaque counts per

dish and the time required for plaque development in 293 cells from three separate experiments. "I" designates I-CeuI; and "B" designates Bsa BI.

FIGURE 4B shows the viral titers obtained from plaques that were allowed to grow over 10 days after transfection. Viruses were harvested and the titer of each virus stock was determined by a GFP based semi-quantitative titration procedure described herein. Values represent the mean ± SE of three independent determinations.

FIGURE 5 is a schematic diagram showing a linear AdV that resolves into a circular episome. The elements involved in the self-directed rearrangement of the vector are shown schematically in pLEP1BHCRGFP/EBV and in the corresponding AdV.

Starting from the left ITR, the elements are shown as following sequence: left ITR, 147 bp; first 34 bp loxP site; 185 bp enhancer/packaging signal; 64 bp splicing acceptor (SA) from EF1 α gene first intron; 720 bp GFP cDNA; 230 bp SV40 poly(A); 1.7 kb TK-EBNA-1/OriP; 970 bp HCR12 promoter; 1 kb EF1 α gene first intron containing splicing donor (SD) and acceptor (SA) sites with the second loxP site inserted at 64 bp upstream of the 3' end; 1.2 kb Cre gene tagged with AU1 and a nuclear localization signal; ~120bp poly(A) signal and PI-PspI site. After infection of liver cells, the HCR12 promoter drives the expression of Cre which results in the cleavage of the two loxP sites. This results in the circularization of the fragment containing the EBV replicon. The excision severs the connection between the enhancer/packaging signals and the remainder of the AdV genome. The Cre gene becomes promoterless and is left on the AdV genome fragment. After excision, the HCR12 promoter drives the expression of the GFP reporter gene. The EBV replicon maintains the excised circle as an episome in host cells.

FIGURE 6A is a schematic representation of the loxP sites and EBNA-1 locations in the AdV genome. The relevant Bgl II site is also shown.

FIGURE 6B shows the time course of rearrangement in HepG2 and Hela cells at an equal multiplicity of infection (moi) of 1,000 particles per cell. Cells were infected with Ad2HCRGFP/EBV viruses for 2 hours at 37 °C. Hirt DNA samples were extracted from the cells. ~5 μ g of Hirt DNA samples were digested with Bgl II, fractionated on a 1% agarose gel, and analyzed by Southern blot techniques using a 32 P-labeled EBNA-1 fragment as the hybridization probe.

FIGURE 6C shows the DNA blot results obtained from Hela cells infected at a moi of 10,000; and HepG2 at 1,000. The upper bands (4915 bp) represent the circularized DNA fragments whereas the lower bands (3162 bp) represent the non-circularized AdV.

- 5 FIGURE 7A shows green fluorescent protein (GFP) expression in liver and non-liver cells infected with the Ad2HCRGFP/EBV viruses. Cells were cultured in 35 mm dishes and infected with the Ad2HCRGFP/EBV virus at desired moi. HepG2 cells were infected with 1,000 particles per cell, whereas Hela cells were infected with a moi of 10,000. GFP expression was examined at the indicated time points after infection.
- 10 10 Fluorescent cells were photographed using an Olympus SC35mm camera mounted on an Olympus IX70 fluorescent microscope, at 200x magnification, using a filter with peak excitation and emission wavelengths of 450 nm and 510 nm, respectively.

- 15 FIGURE 7B shows the expression of GFP in HepG2, Hela, A431, and HT29 cells. Cells were seeded in 35 mm dishes and infected with the Ad2HCRGFP/EBV virus at a moi of 10,000 particles per cell. GFP expression was examined at 72 hours after infection.

FIGURE 7C shows the expression of GFP in human primary hepatocytes. These cells were photographed under bright field (left) and fluorescent conditions (right).

- 20 FIGURE 8A shows the results of RT-PCR that was performed to detect the tripartite leader sequence (upper panel) for virus late gene expression; and PCR was performed in the DNA samples for detection of the AdV genomes. The specific target sequences are described in detail *infra*. PCR analyses of adenovirus late gene expression in cells infected with the first generation AdVs or the self-resolving Ad2HCRGFP/EBV was analyzed. HepG2 cells were cultured in 35 mm dishes and 25 infected with increasing moi (0, 10, 100, 1000, 10,000, and 100,000) of adenoviral vectors. RNA and DNA were isolated in parallel from the cells at 72 hours after infection.

FIGURE 8B shows a summary of quantitative RT-PCR and PCR results. Each determinant was the average of three experiments.

- 30 FIGURE 9A is a schematic diagram depicting the deletion analysis of the OriP and EBNA-1 regions of the EBV replicon. Structures of the deletions in EBNA-1 and OriP are schematically represented. Elements considered important for episomal maintenance are indicated. FR refers to the family of repeats; DS designates the region

of dyad symmetry; LR1 refers to the so-called linker region 1; GA refers to gly-ala repeats; LR2 refers to linker region 2; and Dimerization designates the dimerization domain.

FIGURE 9B is a graph depicting fractions of GFP positive cells carrying the
5 EBV replicons represented in Figure 9A.

FIGURE 10A shows the positions and identities of Cre mutants tested for their dominant negative Cre activities.

FIGURE 10B is a schematic diagram of the substrate Cre plasmid (ad2239) used to test dominant negative functions of Cre mutants.

10 FIGURE 10C shows GFP expression in cells cotransfected with the substrate Cre plasmid (ad2239) and the indicated Cre mutants.

FIGURES 10D and 10E show Cre mutants tested for their ability to inhibit rearrangement. Only those showing the strongest inhibitory activities were retested in Fig. 10E. GFP intensity was normalized to that of cells in the absence of inhibition.

15 FIGURE 11 shows GFP expression in 293 TetON cells and #17 cells transfected with ad2239. The ability of #17 cells to inhibit Cre activity is demonstrated by the weak · GFP signal in cells treated with 2 µM doxycycline.

FIGURE 12 is a schematic diagram depicting the tetracycline mediated auto-regulatory circuit.

20 FIGURES 13A and 13B show the effects of different basal elements on synthetic TetO promoter activity. FIGURE 13A shows a schematic diagram of the components of various auto-regulatory synthetic TetO promoters. FIGURE 13B shows a comparison of the strength of auto-regulatory synthetic TetO promoters bearing different basal elements, in the presence and absence of tetracycline, using GFP as a marker in HepG2
25 cells.

FIGURE 14 shows the structure of a Cre substrate plasmid (ad2265). The promoter, Ef1 α , and the gene, BFP, are interrupted by two loxP sites, which can be joined by Cre-mediated recombination. PA stands for poly A; BFP for blue fluorescent protein.

30 FIGURES 15A and 15B show the estrogen regulation of Cre recombinase activity. 293 cells infected with type B virus, AD121.5, in which the Cre enzyme is fused with estrogen ligand binding domain at both the N- and C-termini were cultured in the presence or absence of 1 µM estrogen. Cre-mediated rearrangement in the presence

of estrogen is shown in Figure 15A, whereas blot analysis of extrachromosomal DNA from the same cells is shown in Figure 15B. L represents the position corresponding to the unarranged adenoviral DNA; and C represents the position corresponding to the circular form of DNA.

5 FIGURE 16 shows the rearrangement of adenoviral sequences *in vivo*. Extrachromosomal DNA from the livers of Rag-2 mice sacrificed 2.5 hrs post injection of type A adenovirus, AD102.7, was analyzed by DNA blot. L represents the size corresponding to linear adenoviral DNA; and C represents the size corresponding to rearranged circular DNA.

10 FIGURE 17 is a photomicrograph depicting high level GFP expression in Rag2 mouse hepatic tissues 48 hrs post type A adenovirus (AD102.7) injection.

FIGURE 18A and 18B show schematic diagrams of the structures of adenoviral vectors and their fates in target cells. enh refers to Ad2 enhancer; GFP refers to green fluorescent protein; EBV refers to Epstein Barr Virus replicon; TetO₇ refers to heptamer 15 of Tet operator; TetR refers to Tet repressor; VP16 refers to transcriptional activator domain from HSV protein 16; SD refers to splice donor site; and SA refers to splice acceptor site.

FIGURE 19A shows the structure of a FLP substrate plasmid, ad2879. The promoter, Ef1 α , and the gene, GFP, are interrupted by 2 FRT sites, which can be joined 20 by the FLP-mediated recombination. PA stands for poly A; BFP for blue fluorescent protein.

FIGURE 19B shows the structure of a cre substrate plasmid, ad2204.

FIGURE 20 shows the structures of several FLPe anti-sense plasmids.

FIGURE 21 is a panel of photomicrographs showing inhibition of FLP enzyme 25 activity by anti-sense FLP. 293 cells were transfected with FLP substrate (Figure 12) and plasmids indicated in each photo. High GFP intensity indicate the higher expression of FLP and less inhibition by the anti-sense expressed.

FIGURE 22 shows a schematic diagram of FRT/Cre and loxP/FLP adenovirus.

Detailed Description

Described herein are systems for the regulated self-rearrangement of DNA vectors, for example, gene therapy vectors. Such regulated self-rearrangement has the potential to prevent unwanted expression of vector genes not required for a therapeutic effect, and to allow the stable association of the therapeutic gene with the target cell.

The essential elements of the regulated DNA rearrangement system are a gene which encodes one or more proteins which induce DNA rearrangement, a method for regulating the activity of those proteins or their abundance, and a target DNA sequence on which those proteins act. Particularly desirable are methods for regulating the activity of the proteins or their abundance which can be easily carried out on an intact organism, such as administration or withdrawal of a drug, hormone, or environmental stimulus such as heat or irradiation, which induces the activity or abundance of the proteins which cause DNA rearrangement.

Especially desirable are regulated DNA rearrangement systems in which all of the components can be delivered in a single vector. An example of this is a virus which bears both the cis-acting sequences for DNA rearrangement as well as the protein or proteins which act on those sequences, and the regulatory apparatus which controls the activity or abundance of those proteins. However, it is not necessary that the different elements be encoded in a single nucleic acid.

The important elements of this strategy are: the compromise of vector gene function by regulated rearrangement of DNA topology, the generation of plasmid circles from vector DNA in a regulated manner, and the removal of enhancer or promoter elements from the vector DNA by regulated excision. It is also important that the circular DNA generated by site-specific recombination possesses a mechanism for stable association with the host genome in some form, here conferred by the EBV replicon. In other embodiments, the circular DNA might possess the ability to direct its integration into the host chromosomes by a site-specific integration. Site-specific integration into the host chromosomes may also be generated by the action of a regulated site-specific recombinase on a linear template without passing through a circular intermediate.

Also described herein is one particular self-rearranging vector that begins as a hybrid adenovirus vector which is capable of converting itself into two unlinked molecules, a circular and a linear DNA. After this event the linear DNA product is deleted for two important cis-acting sequences: the packaging signals, which are

required for insertion of the viral DNA into the viral capsid, and the enhancer, which increases the expression of other promoters encoded in the viral DNA. The remaining linear DNA is thereby compromised in its ability to express adenoviral genes, endowing the vector with a lower immunological profile. The circular DNA generated by the 5 excision event is a mammalian cell plasmid which has the capacity to persist by autonomous replication in the nucleus. This capacity is encoded in genetic elements derived from the Epstein Barr virus (EBV). A schematic diagram of such a vector is illustrated in Figure 5.

Epstein Barr virus is a human herpes virus which is the etiologic agent of 10 infectious mononucleosis and which has been implicated in the genesis of Burkitt's lymphoma, a B cell neoplasm, and is thought to be a predisposing factor for some forms of nasopharyngeal carcinoma. Approximately 85% of the adult Western population has a persistent population of B cells which contain a circular latent form of the viral-genome, maintained in cells by the action of Epstein Barr Nuclear Antigen 1 (EBNA1), 15 a DNA replication protein that acts on the viral latent phase origin of replication, OriP. EBNA1 in and of itself is not thought to promote neoplasia; current thinking places greater weight on the actions of the EBNA2 proteins and LMP, latent membrane protein, in the inception of EBV-associated neoplasm.

Mammalian cell plasmids have been created which bear the EBNA1 gene and 20 OriP. In nonrodent cells, these plasmids persist by replication with each transit of the cell cycle. Multiple transcription units can be borne by these plasmids, allowing regulated expression of diverse gene products.

Preferred adenoviral vectors, shown in Figures 1A and 1B, are linear forms of an 25 EBV plasmid flanked by loxP sites, cis-acting sequences required for site-specific recombination directed by the bacteriophage P1 cre protein. To prepare an adenovirus bearing both the cre protein and loxP sites, it is necessary to insure that the cre protein is not expressed while the vector is being propagated in 293 cells. To lower the immunological profile of the vector, it is also desirable that the cre protein not be expressed after the vector delivered its payload to the target cell and the cre protein 30 performed its function.

To accomplish these objectives, two general approaches have been developed for the production of adenoviral chromosomes that circularize following the regulated expression of site-specific recombinases. In each case, the vector is engineered to allow

for the production of viruses in 293 cells, and to provide transitory expression of recombinase that induces rearrangement in target tissues. The major difference between the two strategies lies in the way the deinduction of recombinase is achieved.

In the first approach, adenoviral vectors are engineered to turn an activating transcription factor into a repressor upon chromosomal rearrangement. Vectors employing this approach are referred to herein as type A vectors (Figure 1A). In the second approach, the recombinase promoter is redirected following chromosomal rearrangement. Vectors utilizing the second approach are referred to as type B vectors (Figure 1B). In both cases a linear chromosome is converted to its circular episomal form and a resulting deleted linear form. The circular DNA contains an Epstein Barr virus (EBV) replicon, which allows synchronous replication of the episome with the host mitotic cycle (Reisman et al., Mol. Cell Biol. 8: 1822-32, 1985; Yates et al., Nature 313: 812-15, 1985). The linear DNA is deleted for the enhancer and E1 genes.

One self-regulated gene switch, employing the type A vector strategy, was designed based on the bacterial transposon Tn10 tetracycline repressor (tetR) gene. In its natural context, the tetR protein binds to specific sequences (tet operator sequences) upstream of a tetracycline resistance gene, preventing transcription of the gene unless tetracycline is present. To adapt this protein for eukaryotic gene regulation, a gene fusion is created between tetR and an active portion of a strong eukaryotic transcriptional activator, the herpes simplex virus VP16 protein. The fusion protein exerts its action on a synthetic promoter created by the insertion of multiple tet operator sequences upstream of a basal promoter element. This configuration allows high-level gene expression whenever the tetR-VP16 fusion protein binds to its cognate operator sequences. Because the tetR protein normally does not bind to its operator in the presence of tetracycline, the activity of this synthetic promoter is high in the absence of tetracycline and low in its presence.

One example of a type A vector is shown in Figure 1A. This self-regulated gene expression cassette, present in a hybrid adenovirus, consists of a bi-directional promoter element in which central tetR binding sites are flanked by divergently oriented basal promoter elements. In one direction the promoter directs the formation of a transcript encoding the cre protein; in the other direction, the promoter directs the formation of a tetR-VP16 fusion protein. The latter differs from the conventional version in bearing a loxP site between the tetR component and the VP16 component. When tetracycline is

present this gene switch is silent. As shown in Figure 1A, upon introduction into a target cell in the absence of tetracycline, the tetR-loxP-VP16 fusion protein is produced, stimulating further production of the fusion protein, and the cre protein. The cre protein then acts to promote site specific recombination between the loxP site in the tetR-loxP-
5 VP16 coding sequences, and a distant loxP site. As a result of this recombination, the fusion protein coding sequence is disrupted so that the promoter no longer directs the formation of a tetR-loxP-VP16 fusion protein, but gives rise to an inert tetR-loxP-VP16 fusion protein for binding to the promoter upstream elements, thereby extinguishing promoter activity.

10 As shown in Figure 1A, the excised circular DNA element contains at least two transcription units. In addition, other transcription units or internal ribosome entry site elements may be used to allow the coexpression of gene products which are useful for extending the persistence of the delivered DNA, regulating expression of the gene of interest, or providing for ablation of the transduced cells once their presence is no longer
15 desirable. In addition, the linear DNA remaining after excision of the circular gene expression plasmid lacks both viral packaging sequences and the cis-acting enhancer. Within this linear DNA, additional loxP sites may be placed to provide for the rearrangement of the remaining vector DNA in the target cell, disrupting the normal topology of the genes, and further thwarting expression.

20 Using the type B vector design strategy, described in greater detail below, a recombinant adenoviral gene delivery system that is capable of undergoing growth phase-dependent site-specific recombination has also been constructed.

The following examples are presented for the purpose of illustrating, not limiting, the invention.

25

TYPE B VECTORS – EXPERIMENTAL RESULTS

Several experimental examples for constructing type B vectors and for carrying out the general approaches of the invention are now described below.

30 Two-Cosmid System for Efficient Construction of Recombinant AdV

To simplify and facilitate the generation of recombinant AdV, a system was established to assemble the desired AdV genome in a single plasmid by ligation (shown in Figures 2A, 2B). The system consists of two component vectors, a left end plasmid,

pLEP, and a right end plasmid, pREP. The left end Ad sequences (nt 1-376) in pLEP include the viral inverted terminal repeat, the cis-acting packaging sequences, and the viral enhancer. Nucleotide (nt) positions described herein refer to the wild type Ad2 sequence in GenBank (J019017). The Ad sequences are followed by the gene

- 5 expression unit intended for delivery and an intron endonuclease (PI-PspI) cleavage site. The right end plasmid contains a PI-PspI site followed by the Ad2 genome from the end of the E1 locus rightward (nt 3527-35937).

pLEP is a small tractable vector for cloning, whereas pREP is much larger and contains less frequently manipulated genes. Both pLEP and pREP contain a
10 bacteriophage λ cos site, oriented to generate a single cosmid of appropriate length for *in vitro* packaging following ligation of the two plasmids at the PI-PspI cleavage site. pLEP is tetracycline resistant (Tet^r) and pREP is ampicillin (Amp^r) resistant, allowing the recombinants to be selectively isolated by co-selection for both markers. In the resulting assembled cosmid, the adenoviral sequences are closely flanked by cleavage
15 sites for the intron endonuclease I-CeuI. Digestion with I-CeuI liberates the entire recombinant AdV genome from the parent cosmid (see Figure 2B).

Three classes of pREP have been constructed to allow the preparation of AdVs bearing E1 (pREP7; SEQ ID NO.: 2), E1 and E3 (pREP8; SEQ ID NO.: 3), or E1, E3, and E4 (pREP12; SEQ ID NO.: 4) deletions. pREP7 (SEQ ID NO.: 2) contains nt
20 3527-35937 of the Ad2 genome, and pREP8 (SEQ ID NO.: 3) carries an additional deletion in the E3 region (Δ nt 27901-30841). pREP12 (SEQ ID NO.: 4) has deleted open reading frames (ORF) 1-4 of the E4 region (Δ nt 34121-35469, 1348 bp). AdV generated with these cosmids should be able to accommodate 5, 8, and 10 kb inserts, respectively.

25 These aforementioned vectors were constructed as follows. The EcoRI to BsaI fragment that spans the ampicillin resistance gene in pBR322 was deleted and replaced by a synthetic adapter, and the bacteriophage λ cos site was inserted between the unique StyI and BsmI sites. A PCR amplified Ad2 fragment containing the left end ITR (L.ITR), enhancer elements, and the encapsidation signal (nt 1-376) was created and
30 inserted into the adapter (Figures 2A, 2B) to yield the tetracycline-resistant left-end plasmid pLEP. The right end of Ad2 from the AflII site to the right end (nt 3527-35937) was assembled into an ampicillin resistant cosmid vector, pACKrr3 (SEQ ID NO.: 1), by multiple steps of PCR amplification and fragment interchange. The resultant cosmid

was termed pREP7 (SEQ ID NO.: 2). To expand vector capacity, two deletions were incorporated into the pREP7 (SEQ ID NO.: 2) cosmid, an E3 gene deletion (nt 27901-30841, 2840 bp); cosmid pREP8 (SEQ ID NO.: 3) and a 1.3 kb deletion (nt 34121-35469) in the E4 region of the Ad2 region; pREP12 (SEQ ID NO.: 4)

5 An example of the construction of an AdV carrying a CMV-GFP expression unit is outlined in Figure 2. pLEPCMVGFP (Tet^r) was digested with PI-PspI and ligated to the pREP7 (SEQ ID NO.: 2; ΔE1, Amp^r) digested with the same enzyme. The ligation mixture was packaged with λ phage extracts (MaxPlax lambda packaging extracts, Epicentre Technologies) and a fraction of the packaged phage was used to infect a
10 recombination-deficient *E. coli* host, with selection for the assembled plasmid on Amp/Tet plates. Transductants containing pLEP fused to pREP were selected on agar containing 25 μg/ml ampicillin and 12.5 μg/ml tetracycline (Amp/Tet). Colonies were selected and DNA isolated (Qiagen). DNA was used either for restriction analysis or for transfection of 293 cells as described herein.

15 Figure 3A shows typical results for the Bgl II digestion pattern of a pLEP3CMVGFP/pREP7 hybrid cosmid, pAd2-7CMVGFP DNA. Because of the size minimum (~40 kbp) for λ phage *in vitro* packaging and the double antibiotic selection, most of the colonies growing on Amp/Tet plates were the desired hybrid cosmids, and undesired rearrangements were rarely seen. In the present example, all four pAd2-
20 7CMVGFP clones exhibited the digestion pattern predicted from the inferred sequence. The entire recombinant AdV genome was then released from the cosmid by I-CeuI digestion (Figure 3B). I-CeuI digestion leaves ten nucleotides to the left of the left ITR and eight nucleotides to the right of the right ITR. Short flanking sequences have been reported to be eliminated during replication of recombinant viruses after transfecting the
25 DNA into 293 (human embryonic kidney) cells (Hanahan et al., Mol. Cell. Biol. 4:302-309, 1984).

The digestion reaction can be transfected into 293 cells without purification as follows. 293 cells, obtained from Microbix Biosystems (Ontario, Canada), were cultured in 10 cm dishes in complete Dulbecco's Modified Eagle's Medium (DMEM)
30 supplemented with 10% FBS, 2mM glutamine and penicillin/streptomycin (Gibco BRL), and maintained at 37 °C and 5% CO₂ atmosphere incubator. Cells were grown to ~50% confluence on the day of transfection. Ten μg of cosmid DNA were digested with I-CeuI in a volume of 50 μl. The reaction mixture was transfected into 293 cells by

calcium phosphate precipitation (Graham and Prevec, Manipulation of adenovirus vectors, p. 109-128, *In E. J. Murray (ed.), Methods in Molecular Biology*, vol. 7, Humana, Clifton, NJ, 1991) without purification. After transfection, cells were cultured and examined daily for the appearance of cytopathic effects (CPE). Virus propagation, 5 purification, plaque assay, and viral DNA isolation were performed using established protocols (Graham and Prevec, *supra*). At day six post-transfection, 5-30 viral plaques/10 cm dish /10 µg DNA were usually apparent, which compared favorably with the 30-50 plaques/10 cm dish /10 µg DNA found for 293 cells transfected with purified wild type Ad2 DNA.

10 To compare the efficiency of recombinant virus production, similar viruses were also generated by homologous recombination. 20 µg of pREP7 (SEQ ID NO.: 2) was co-transfected into 293 cells with 10 µg of a plasmid encoding the left end of the adenoviral genome and a green fluorescent reporter gene (pLITREF1αGFP).

pLITREF1αGFP contained the Ad2 left end nt 1-376, an EF1α promoter/GFP expression unit and Ad2 sequence (from 3525-8120) that overlaps with the same sequence in pREP7 (SEQ ID NO.: 2). This overlap fragment served as the region for homologous recombination. Each co-transfection was performed in duplicate. Initial plaques took longer to appear (14 days post transfection) and were less abundant (0-3 plaques per plate).

20 Data in the literature suggest that exposed ITR ends favor efficient virus production (Hanahan et al., *supra*). To assess the importance of this effect, an AdV cosmid, pIAdEF1αGFPB, in which the AdV ITRs were flanked with a different restriction site at each end was constructed. pIAdEF1αGFPB DNA was digested with BsaBI to expose the right ITR, I-CeuI to expose the left ITR, or the two enzymes were 25 used together to expose both ends. Digested cosmid DNA samples were transfected into 293 cells and plaques were allowed to develop. Virus propagation, purification, plaque assay, and viral DNA isolation were performed using established protocols described in Graham and Prevec. (*Manipulation of adenovirus vectors, In E. J. Murray (ed.), Methods in Molecular Biology*, vol. 7. Humana, Clifton, NJ., pp. 109-128, 1991).

30 Ten days after transfection the viruses were harvested and viral titers were determined. The average titer for the viral stocks (Figures 4A, 4B) was 1.3×10^4 pfu/ml from transfection with undigested DNA; 2.4×10^5 pfu/ml from BsaBI linearized DNA (free right ITR); 1.1×10^5 pfu/ml from I-CeuI linearized DNA (free left ITR); and $2.7 \times$

10^6 pfu/ml for the BsaBI/I-CeuI double digested DNA (both ITRs free). Thus liberation of each end resulted in an approximate increase in the efficiency of generating virus by a factor of ten (Figures 4A, 4B).

5 Construction of an AdV Capable of Self-Rearrangement

One approach to attenuating adenoviral gene expression and improving transgene persistence is the creation of viruses capable of undergoing internal, self-directed rearrangement upon delivery to the target tissue. In principle, this objective can be achieved through the regulated expression of site-specific recombinases in vectors that contain the cis-acting target of recombinase action. To allow such vectors to be created, the recombinase activity must be suppressed during propagation in the packaging cell line. As described in more detail below, the use of a lineage-specific promoter to control recombinase expression has been successfully employed to achieve this end.

An example of this is shown in Figure 5. The expression of Cre recombinase was controlled by a liver-specific promoter constructed as follows. The human hepatic control region 1 and 2 (HCR1 and 2) of the ApoE/C gene locus (Allan et al., J. Biol. Chem. 270:26278-81, 1995; and Dang et al., J. Biol. Chem. 270:22577-85, 1995) were amplified by PCR using 293 cell genomic DNA as the template. The following primers were used to amplify both HCR1 and HCR2 fragment: HCRtop-
5' gcggattcggctggacttagagaacagag 3' (SEQ ID NO.:5); HCRbot - 5'
gcgggatcctgaacccggacccttcacacta 3' (SEQ ID NO.:6). The amplified PCR fragments (~0.39 kb) were cloned into pUC19. The HCR1 and HCR2 sequences were confirmed by dideoxy DNA sequencing. The two fragments were assembled in a head to tail orientation, fused with a synthetic basal TATA element and cloned in a parental pLEP vector containing a GFP reporter gene. The resultant plasmid was named pLEPHCR12GFP. The synthetic liver-specific, as demonstrated below, provided a means to control Cre recombinase expression during propagation of the vector in 293 cells, and allowed for testing the consequences of abstracting the enhancer from the linear vector DNA upon delivery of the DNA to the target cells.

30 In 293 cells, this promoter is silent, allowing the viral chromosome to be propagated with minimal rearrangement. Any rearranged viruses that are formed lack packaging signals and so disappear from the pool of propagating vectors. In liver cells the Cre recombinase is induced by the action of the tissue-specific promoter. The

resulting Cre-induced recombination excises a circular episome and redirects the transcriptional output of the liver-specific promoter so that it directs the synthesis of the transgene of interest. The remaining linear fragment consists of an adenoviral genome lacking the enhancer and packaging signals and a Cre expression unit devoid of
5 promoter sequences.

In the form discussed here, one loxP site is located at nucleotide 147 of the Ad2 genome, between the left ITR and the enhancer/packaging sequences, and the second loxP site is placed inside an intron a few bases upstream of the splice acceptor sequence. Hence the loxP site does not appear in the resulting mature transcript. The Cre coding
10 sequence that remains on the right end linear fragment after rearrangement lies downstream from a splice acceptor that lacks a splice donor or upstream promoter sequences. This effectively terminates the expression of Cre following excision.

Prior to recombination, the Cre recombinase gene is under the control of a synthetic promoter (referred to as HCR12), consisting of hepatic locus control elements
15 from the human ApoE/C locus fused to the first intron of the human EF1 α gene. After cyclization the HCR12 promoter lies upstream of the transgene (in this case GFP) and the distal segment of the intron (beyond the loxP site) contains the adenoviral enhancer. To facilitate manipulation of the plasmids in *E. coli*, the human IgG1 hinge-CH2 intron
20 (118 bp) was inserted in the Cre coding sequence at nucleotide 237, suppressing Cre expression in bacteria. The circularized episome contains the latent origin of replication (OriP) and trans-acting DNA replication protein (EBNA-1) of Epstein Barr virus, and hence is capable of autonomous replication in synchrony with the host mitotic cycle
(Yates et al., Nature 313:812-815, 1985).

Using the two cosmid system described above, the pLEP plasmid containing the
25 self-resolving components, pLEP1BHCR12, was ligated with pREP8 (SEQ ID NO.: 3; Δ E1 Δ E3) to create pAdVHCRGFP/EBV. The latter was digested with I-CeuI and transfected into 293 cells. Appearance of plaques from AdVHCRGFP/EBV was retarded (by 8 days) compared to non-rearranging viruses, perhaps as a result of basal expression of the liver-specific promoter in 293 cells. However high titer viral stocks of
30 10^{12} nominal (absorbance-determined) particles/ml was achieved.

Rearrangement in Target and Nontarget Cells

To test excision efficiency, HepG2 (hepatocellular carcinoma) and Hela (cervical carcinoma) cells, obtained from ATCC, were infected with virus at a multiplicity of infection (moi) of 1,000 nominal particles/cell. This titer corresponds to approximately 5 10 plaque forming units per cell. For these experiments, HepG2 and Hela cells were seeded in 35mm dishes and cultured to approximately 80% confluence in DMEM/FBS as described herein. Cells were infected with the desired multiplicity of virus in a volume of 1 ml at 37 °C for 2 hours. At the end of the incubation, cells were washed with PBS twice and cultured in 2 ml of medium. Cells were collected in parallel at 10 desired points for low molecular weight DNA and RNA extraction. Cells were examined for GFP expression by fluorescence microscopy (Olympus, IX70) or microtiter plate reader (PerSeptive Biosystem, CytoFluor II) before extraction of DNA for analysis of chromosomal rearrangement.

DNA analysis of chromosomal rearrangement was performed as follows. 5 µg 15 of Hirt DNA was digested with Bgl II and analyzed by DNA blot techniques using a labeled EBNA-1 gene fragment as probe (Figure 6). The Bgl II fragment from the non-circularized AdV is 3162 bp, generated from the 5'end of the AdV to the first BglII site in the AdV. The circularized fragment created from the two loxP sites has a size of 4915 bp (Figure 6A). Densitometry revealed that at 72 hours post infection, 95% or 20 more of the input genomes had undergone circularization in HepG2 cells. In contrast, low but detectable levels of circularized fragment was visualized in Hela cells infected at the same time and at the same multiplicity of infection used for the HepG2 cells (Figure 6B).

At the time of infection (t=0, Figure 6B), the amount of input viral DNA detected 25 by DNA blot was higher for HepG2 cells than for Hela cells when similar virus multiplicities were applied (moi of 1,000). This may reflect differences in AdV adsorption or infection efficiency between the two cell types, possibly as a result of the lower levels of coxsackievirus-adenovirus receptor on the Hela cells surface. To achieve similar viral genome input into HepG2 and Hela cells, Hela cells were infected with ten-fold more virus (moi of ~10,000) than HepG2 cells (moi of 1,000). Episomal DNA samples were extracted and analyzed by blotting. The results (Figure 6C) indicated that when comparable amounts of viral genome are present in the nucleus, the cyclization rate in both cell types was similar. Because the level of subsequent GFP expression is

much higher in HepG2 cells than in HeLa cells (Figure 7A), it is likely that very small amounts of Cre recombinase suffice to promote rearrangement, and that recombinase expression is not limiting for rearrangement in either HepG2 or HeLa cells.

GFP expression cannot be detected until rearrangement has taken place, so the measurement of the fraction of GFP positive cells provided a simple alternate method for assessing the degree of productive rearrangement. Figure 7A shows that GFP expression developed quickly in transduced HepG2 cells, but that only a few GFP positive cells can be detected in Hela cells infected with a ten fold higher moi, conditions that allow circularization to a comparable extent to that seen in HepG2 cells (Figure 6C).

The HCR12 promoter specificity was also tested by infecting two additional non-hepatic cell lines, A431 (human epidermoid carcinoma) and HT29 (human colon adenocarcinoma), with the Ad2HCRGFP/EBV vector. Both cell lines were obtained from ATCC and cultured using DMEM/FBS as described herein. A few cells, with weak GFP signal, were detected at 72 hours after infection in these cells (Figure 7B). In contrast, these non-hepatic cells could be infected efficiently with a first generation AdV, Ad2CMVGFP virus (data not shown), indicating that the low GFP signal was not due to the low infectivity of these cells by AdV.

To further assess the utility of the AdV genome rearrangement, primary human hepatocytes were infected with the Ad2HCRGFP/EBV vector. For these experiments, primary human hepatocytes, generously provided by Dr. Albert Edge (Diacrin, Inc., Charlestown, MA) were isolated and cultured as described by Gunsalus et al. (Nat. Med. 3:48-53, 1997), infected with adenovirus, and GFP expression was analyzed. As shown in Figure 7C, GFP expression was readily detected 72 hours after infection.

25

Diminished Viral Gene Expression in Rearranged AdV

After excision, the adenovirus major enhancer/packaging signal segregates with the episomal DNA, yielding a linear fragment containing the remainder of the AdV genome without this important cis-element (Figure 5). To assess the impact of enhancer deletion, PCR amplification and quantitative RT-PCR measurement of late viral gene expression was performed as follows.

Four µg of total RNA was reverse transcribed into cDNA using M-MLV RT by a standard protocol (Promega). 1 µl of the cDNA from each sample was used in subsequent PCR reactions. PCR primers were designed to amplify the tripartite leader sequence of the adenovirus late genes: TPL1 - 5' act ctc ttc cgc atc gct gt 3' (SEQ ID NO.: 7) and TPL2 - 5' ctt gcg act gtg act ggt tag 3' (SEQ ID NO.:8). For detection of the AdV genome in the Hirt DNA samples, 1 µg DNA was employed in the PCR amplification using the following primers which are specific for the adenovirus DNA in the fiber gene: Fiber1 - 5' ccg cac cca cta tct tca ta 3' (SEQ ID NO.: 9) and Fiber2- 5' ggt gtc caa agg ttc gga ga 3' (SEQ ID NO.: 10). PCR reactions were performed as 95 °C 30 seconds; 54 °C 30 seconds; 72 °C 30 seconds for 30 cycles. All amplified products were analyzed on a 2% agarose gel.

For quantitative PCR, a molecular beacon based universal amplification and detection system was used (Intergen). A common leading sequence (Z sequence, 5' act gaa cct gac cgt aca 3') was added to the TPL1 and Fiber1 primers. The TPL2 and Fiber2 primers, described above, were used in the quantitative PCR reactions. 1 µl of the cDNA and one µg of Hirt DNA from each sample were used in the assay. The PCR were carried out in a 96-well spectrofluorometric thermal cycler (Applied Biosystems Prism 7700). The number of template molecules in the PCR reaction was calculated from the standard curve using linearized plasmid as templates.

As most late adenoviral genes transcripts share a common ~200 bp tripartite leader sequence (TPL) (Akusjarvi and Persson, Nature 292:420-6, 1981), the TPL sequence was chosen as a marker of viral gene expression. HepG2 cells were infected with the first generation vectors Ad2CMVGFP and Ad2HCRGFP, or the self-resolving vector, Ad2HCRGFP/EBV, using increasing multiplicities of infection. Total cellular RNA and low molecular weight DNA were isolated in parallel as described by Hirt (J. Mol. Biol. 26:365-9, 1967) and total RNA was prepared using RNAzol solution (Tel-Test. Inc.). RT-PCR was performed to quantitate the amount of RNA encoding the TPL in the cDNA samples. PCR amplification of a 201 bp fiber gene fragment from the AdV genome was used to detect the amount of viral genome in the DNA samples. A representative result of three experiments is shown in Figure 8A. TPL sequences were detected, 72 hours post-infection, with either 100 or 1000 viruses infected per cell, using both of the first generation adenoviruses (upper panel).

In contrast, no TPL signal was detected in the self-resolving Ad2HCRGFP/EBV infected cells, even at a moi of 100,000/cell. PCR amplification of the AdV fiber gene revealed comparable levels of AdV genomic DNA in cells infected at comparable moi's. (Figure 8A, lower panel). The cDNA samples in which the TPL signals were detected 5 were further analyzed by real-time fluorescence PCR. The corresponding genomic DNA samples were also analyzed to determine the number of AdV genomes present in each sample. The results are summarized in Figure 8B. There were approximately 1×10^4 TPL per 1×10^6 AdV genomes detected in the Ad2HCRGFP infected cells, but no detectable TPL in the self-resolving Ad2HCRGFP/EBV infected cells. These results 10 indicate that adenoviral gene expression was dramatically reduced by the separation of the viral enhancer sequences occasioned by the re-arrangement of the self-resolving vector.

TYPE A AND TYPE B VECTORS – EXPERIMENTAL RESULTS

15 Additional experimental examples now follow that further illustrate the general approaches of the invention relating to using and constructing type A and type B vectors. For generating such adenoviral vectors, DNA sequences important for gene expression in the target tissue were placed between two loxP sites. The first loxP site was inserted between the Ad2 left-end inverted terminal repeat (ITR) and the enhancer sequence, 20 replacing a BspLU11I and BstZ17 fragment of Ad2. A target gene expression cassette, comprising a promoter, a gene of interest, polyadenylation signals, the EBV replicon, and site specific recombinase expression unit were inserted in place of the E1 locus.

In type A adenoviral vectors, the second loxP site is placed between TetR and VP16, preserving the coding frame of both (Figure 1A). A bidirectional promoter in 25 which a central heptamer of tetracycline operator sites (TetO) (Gossen and Bujard, Proc. Natl. Acad. Sci. USA 89:5547-5551, 1992) was flanked by two divergently oriented basal elements, directs the expression of TetR loxP VP16 from a synthetic TATA element, whereas Cre recombinase is controlled by the same heptamer of operator upstream of the HIV LTR basal element.

30 In the case of type B viruses (Figure 1B), the second loxP site was inserted in the first intron of the Ef1 α gene, which contains the transcription stimulating sequences described herein. In addition, a splice acceptor sequence was added to the 5' end of the coding sequence of the gene of interest. To avoid rearrangement during plasmid

construction in bacteria, the Cre recombinase coding sequence was interrupted by the addition of the human IgG1 hinge-CH2 intron (between amino acids Q78 and A79), as described herein.

5 Designing a Compact EBV Replicon

Most plasmids employing the EBV latent origin of replication exceed 10 kb in length. To provide a means for increasing the capacity of the recombinant adenoviral type A or type B vectors to accommodate a therapeutic gene, a compact EBV replicon having episomal stability was designed. To this end, deletions were generated in both 10 the cis-acting origin of replication, OriP, and the sequences encoding the trans-acting replication protein, Epstein Barr virus nuclear antigen-1 (EBNA-1) (Figure 9A). Episomal persistence was assessed with a green fluorescent protein (GFP)-bearing test plasmid by determining the fraction of cells retaining green fluorescence as a function of time, assuming that the half-life of GFP, in daughter cells that have not received an 15 episome as a result of segregation failure, is approximately 1.4 days (Fukumura et al., Cell 94:715-725, 1998).

EBNA-1 contains a central repeated structure that consists entirely of Gly and Ala residues, termed the GA repeats (Figure 9A). Although deletion of this structure has been reported to have little consequence, a deletion mutant consisting of both a short 20 OriP and a short EBNA-1 (SoriP + SEBNA1) was generated and found not to support plasmid maintenance effectively (approximately 40% loss per cell division). A version of this mutant, reconstructed with 40 GA repeats, in which the short OriP was paired with a short EBNA-1 provided significantly better plasmid stability (20% loss per cell division vs. 10 % per cell division for the wild type) (Figure 9B). Since most target 25 tissues are relatively quiescent mitotically, this level of segregation fidelity provides reasonable stability in a compact replicon.

Producing Cell Lines that Express Cre- or FLP-Dominant Negative Mutants

As discussed herein, one obstacle to creating adenovirus carrying both 30 recombinase and target sites has been the difficulty of controlling recombinase activity during virus propagation. Since efficient recombinase activity is needed in target cells, recombinase activity is best tempered in the production cell line.

Vector-independent methods to suppress recombinase activity during the production phase are attractive because they allow vector design objectives to be pursued with fewer constraints. In principle, dominant negative recombinase mutants provide the desired antagonism of recombinase activity. Cell lines expressing such recombinase dominant negative mutants were produced as follows.

5 Dominant negative Cre mutants were selected from known point mutants (Wierzbicki et al., J. Mol. Biol. 195:785-794, 1987) that are defective in recombination function but are likely to retain dimerization function Figures 10A-E. Several mutants were screened for their abilities to inhibit Cre activity of a type B vector construct
10 (ad2239 in Figure 10B) in a transient cotransfection assay. Under these conditions, Cre activity is detected by the expression of GFP that occurs upon rearrangement. Figures 10D and 10E show the point mutants that were assessed and their relative activities in the transient cotransfection assay. Dilution studies, in which increasing amounts of substrate/Cre plasmid were cotransfected with the mutant forms, were conducted and
15 based on its favorable profile, one mutant recombinase, designated CreY324C, was chosen for further development (Figure 10D).

Strong constitutive expression of CreY324C, under control of the Ef1 α promoter failed to yield stable cell lines. Stable clones were obtained when the Ef1 α promoter was replaced with a tetracycline regulated promoter (Gossen et al., Science 268:1766-
20 1769, 1995). Clones were then tested for the ability to inhibit Cre enzymatic activities, and one clone, designated cell line #17, was selected for additional experiments. When a plasmid bearing Cre and capable of undergoing Cre-directed rearrangement to create a GFP transcription unit (ad2239) was transfected to #17 cells or parental 293ON cells,
25 GFP expression in the #17 cells in the presence of 2 μ M doxycycline was significantly lower than those of controls (Figure 11), showing that Cre enzyme activity can be inhibited in #17 cells.

In addition to dominant negative Cre mutants, dominant negative FLP mutants may also be identified. FLP belongs to the same family of site-specific recombinases as Cre recombinase. A number of FLP mutations that show defects in either cleavage or
30 ligation of FRT sites have been identified. Mutant FLP defective in cleaving FRT site (for example, H309L, L315P, G328R, G28E, N329D, S336Y, S336F, A339D, Y343F, and H345L) are generated using standard methods. Mutants that inhibit the wild type enzyme are then identified for generating stable cell lines according to the methods

described above. These and the other cell lines (described herein) are then used for producing FRT/FLP containing virus.

As mentioned above, difficulties creating stable cell lines expressing Cre dominant negative mutants were occasionally encountered. This difficulty is not limited
 5 to Cre mutants, but also to the wild-type Cre enzyme. In contrast, 293 cell lines stably expressing a thermostable FLP, referred to as FLPe (Buckholz et al., Nat. Biotechnol. 16:657-662, 1998), were created, suggesting that FLPe might not be as cytostatic as Cre protein. To demonstrate this, 293 cells were transfected with plasmids expressing either Cre or FLPe, and puromycin resistant colonies were selected. To generate stable cell
 10 lines expressing Cre or FLP mutants, 293 TetON cells were transfected with linearized plasmid expressing Cre or FLP mutants and puromycin acetyltransferase and selected with 1 µg/ml of puromycin. Puromycin resistant colonies were characterized further for their ability to inhibit Cre recombinase using the cre (ad2239) or flp (ad2879) substrate plasmids. Table 1 shows that there are more puromycin resistant colonies selected from
 15 FLPe transfected cells than from Cre transfected cells. From this result, it is expected that stable cell lines expressing a reasonably high level of dominant negative FLP may be readily created.

TABLE 1

20

Puromycin Resistant Colonies Formed When Cre Expressing or FLPe Expressing Plasmid was Used to Transfect 293 Cells

25

| Expression Plasmid | Number of colonies (2 µg/ml puromycin) |
|--------------------|--|
| Control | 236 |
| Cre | 92 |
| FLPe | 127 |

25

Cre or Cre dominant negative mutants were also found to inhibit FLP activity (Table 2). Accordingly, cell lines such as cell line #17, that stably express a Cre dominant negative mutant (for example, CreY324C), are useful for producing FLP/FRT 30 carrying adenovirus.

TABLE 2

Cre Inhibition of FLP Activity *in trans*

| Plasmids | Arbitrary GFP intensity |
|---|-------------------------|
| Ef1 α FLP + FLP substrate + vector control | 4.9 |
| Ef1 α FLP + FLP substrate + Ef1 α Cre | 0.78 |
| Ef1 α FLP + FLP substrate + Ef1 α Cre R173C | 2.23 |

5

FLP enzyme activity was measured by the GFP intensity by cotransfected with a FLP substrate plasmid, ad2879 (Figure 19A). GFP intensity was quantified using IP lab software.

10 Transcriptional Regulation of Cre or FLP Recombinases

It has been relatively difficult to achieve high-level promoter inducibility in a replicating adenovirus. The challenge is similar to that of achieving faithful control of transcription in a transient expression setting. One approach to increase the induction ratio in a transient setting is the use of auto-regulatory (feed-forward) circuits. One such system, based on tetracycline dependent activation, is shown in Figure 12. A central heptamer of tetracycline promoter operator sites (TetO sites) was placed between two divergently oriented basal TATA elements. The leftward TATA controls the expression of the TetR-VP16 fusion protein, in which a loxP (or FRT) site has been placed between the TetR DNA binding domain and the VP16 transcriptional activator. The rightward TATA box directs the synthesis of recombinase, either Cre or the yeast FLP enzyme. In the presence of tetracycline, the promoter has reduced activity in both directions. Upon removal of tetracycline, the synthesis of both TetR-VP16 and recombinase are induced (Figure 1A). The induced recombinase then disjoins the TetR DNA binding element from the transcriptional activation contributed by VP16. Any existing TetR-VP16 fusions thereafter promote transcription of TetR, which competes with TetR-VP16 for TetO, resulting in deinduction of recombinase transcription.

When a model target cell line, HepG2, was tested with this type of adenovirus, the efficiency of circularization was low relative to that seen in 293 cells (data not shown), indicating a cell dependence of the bidirectional TetO promoter. To correct this, the TATA element of the TetO synthetic promoter (derived from the CMV immediate early promoter) was replaced with that of the HIV LTR. Constructs bearing differing components of the HIV basal promoter were analyzed for strength and regulation in 293 and HepG2 cells (Figure 13A). Among the constructs tested, one

version bearing the HIV LTR TATA and Sp1 elements (D in Figure 13A) showed the least basal expression in 293 cells (data not shown) and the greatest induction in HepG2 cells (Figure 13B).

Using this promoter, a construct (ad3400) containing the autoregulatory structure D as shown in Figure 13A was engineered, and Cre activities in the presence and in the absence of tetracycline were assayed. Plasmid ad2265 (Figure 14) in which a blue fluorescent protein (BFP) expression unit is interrupted by two loxP sites and transcription termination sequences was used as a substrate for Cre. Cre-mediated recombination joins BFP to the promoter resulting in BFP expression. As shown in Table 3, no difference was found in the intensity of BFP expression, either in the presence or absence of tetracycline. One possible explanation for this is that very little Cre protein is required for activity. Consistent with this idea, standard immunohistochemical techniques failed to reveal the presence of Cre enzyme in cells that were fully induced (data not shown).

15

TABLE 3

20

Cre Recombinase Activity Regulation in Type A Constructs

| Construct | Cre Form | +tet -tam | -tet -tam | +tet +tam | -tet +tam |
|-----------|-------------|-----------|-----------|-----------|-----------|
| ad3400 | Cre | 2.58 | 3.39 | 2.71 | 6.20 |
| ad4394 | Cre-LBD | 0.081 | 0.22 | 0.056 | 1.02 |
| ad4705 | LBD-Cre-LBD | ND | ND | ND | ND |

25

Cre enzyme activity was measured in the presence or absence of the ligands, tamoxifen, by cotransfected with the substrate plasmid (ad2265). BFP intensity (mean intensity/area) was quantified by analyzing fluorescent images captured by a digital camera using IP lab software. ND, refers to fluorescent intensities that were too weak to measure. Tet, Tetracycline; tam, tamoxifen.

30

Deletion of the PolyA Consensus Sequence from Cre or FLP Transcription Units

To reduce the expression of FLP or Cre recombinase further, the consensus polyA addition signals from the Cre or FLP transcript unit were deleted from vector constructs, leaving polyadenylation dependent on distal downstream sequences, for example, in gene IX. The activity of Cre using type B proviral constructs with or without the polyA signal was measured. As shown Table 4, the construct without polyA signals (AD229.3) showed a significant reduction of GFP intensity compared to a

construct bearing the polyA signal (AD230.5). When FLPe constructs of similar structure were evaluated, similar results were found (data not shown). These data show that Cre and FLPe enzyme activity levels can be modulated by attenuating polyadenylation.

5

TABLE 4

| <u>Effect of Deleting polyA Addition Signal From the Cre Expression Unit on Cre Enzyme Activity Level</u> | | |
|---|-------|------------------------|
| 10 | polyA | Relative GFP Intensity |
| AD229.3 | - | 0.25 |
| AD230.5 | + | 1 |

Post-Transcriptional Regulation of Cre Recombinase Activity

Post-transcriptional control mechanisms of Cre recombinase activity were also evaluated. Translational fusions between Cre and the ligand binding domain (LBD) of estrogen receptor have been reported to be regulated by estrogen (Feil et al., Proc. Natl. Acad. Sci., U.S.A 93:10887-10890, 1996; Gossen et al., Proc. Natl. Acad. Sci., U.S.A. 89:5547-5551, 1994), or, in the case of mutant estrogen receptors (Metzger et al., Proc. Natl. Acad. Sci. U.S.A. 92:6991-6995, 1995), by the partial antagonist tamoxifen.

20 Use of a ligand-dependent recombinase (ad4394 in Table 3), in combination with the HIV LTR-based autoregulated Tet system, allowed for a small degree of regulation by tetracycline, but not by ligand, as assayed using the ad2265 rearrangement assay (Table 3). One interpretation of this finding is that fusion of the estrogen receptor LBD to Cre provides only modest control of recombinase activity, but attenuates enzyme 25 potency to a level so that transcriptional regulation can be measured.

To increase control of recombinase activity, the LBD was fused both to the N-terminus and C-terminus of Cre (LBD-Cre-LBD) and inserted into the coding sequence of both type A and type B vectors. When the LBD-Cre-LBD construct of type A was transfected into 293 cells, it showed no significant Cre enzyme activity even in the 30 presence of ligand (Table 3). This result confirmed that the Cre recombinase activity is attenuated by N-terminal or C-terminal extension.

When the LBD fusion Cre enzymes were assayed in the type B vector context, only LBD-Cre-LBD fusions (pk8-ad4626) showed ligand-dependent regulation of Cre enzyme activities (Table 5). It appears that attenuated Cre activity in LBD-Cre-LBD, in the absence of ligand, is low enough to fall below the upper limit of the Cre assay.

5

TABLE 5

Cre Enzyme Activities of Type B Proivirus

| Cre Form | Proivirus | -tam | +tam |
|-------------|------------|------|------|
| Cre | pk8-ad2239 | ND | ND |
| Cre-LBD | pk8-ad4332 | 4.1 | 6 |
| LBD-Cre-LBD | pk8-ad4626 | 0.05 | 4.2 |

10

Cre enzyme activity was measured in the presence or absence of the ligands, tamoxifen. GFP intensity was quantified using IP lab software.
tam, tamoxifen.

15

Consistent with this notion, only the construct carrying two LBDs, pk8- ad4626, was able to produce virus (AD121.5) by transfection and propagate in 293 cells, while pk8-ad4332, which carried one LBD, produced virus (AD100.9) initially (following transfection of the cognate DNA) but was unable to propagate in 293 cells (Table 6). In 20 the case of wild type Cre, no virus was produced in 293 cells by transfection.

20

TABLE 6

Production and Propagation of Type B Adenovirus

25

| Cre Form | Type B adenovirus | Viral Production in 293 cells | Viral Propagation in 293 cells | Viral Propagation in #17 cells |
|-------------|-------------------------|-------------------------------|--------------------------------|--------------------------------|
| Cre | Pack8-2239 | - | - | - |
| Cre-LBD | AD100.9 (Pack8-4332) | + | - | + |
| LBD-Cre-LBD | AD121.5 (Pack8-4626) | + | + | + |

30

The AD100.9 virus was able to propagate in #17 cells expressing the dominant negative Cre Y324C, demonstrating that modulation of Cre activity is important for viral production. Thus, adenovirus carrying both two loxP sites and Cre in two different configurations were generated by controlling Cre activity.

5

Viral Rearrangement in Culture

Cre/loxP mediated rearrangement of the adenovirus in tissue culture cells has also been analyzed. As shown in Figure 15A, the AD121.5 virus showed a significant increase in GFP expression in the presence of the ligand, estrogen, suggesting a 10 successful rearrangement of the virus by Cre recombinases. When non-chromosomal DNA (Hirt, *J. Mol. Biol.* 40:141-144, 1969) was made from the cells and analyzed by DNA blot analysis, the viral DNA from estrogen treated cells was identified mostly in circular form (C in Figure 15B), while the DNA from cells not treated with estrogen was found mainly in linear form (L in Figure 15B).

15

To evaluate the efficiency of the self-rearranging viruses *in vivo*, high titer stocks of AD102.7 (a type A virus carrying LBD-Cre, pk8-ad4394) in #17 cells was prepared and purified by CsCl gradient ultracentrifugation. The titer of AD102.7 ($4 - 6 \times 10^{12}/\text{ml}$ by OD) is comparable to or slightly exceeds that of control viruses ($2 - 4 \times 10^{12}/\text{ml}$ by OD) which carry neither Cre nor a loxP site. To determine the efficiency of viral 20 rearrangement *in vivo* and whether such rearrangement is dependent on the presence of ligand, AD102.7 virus (4×10^{11} pfu/mouse as determined by optical density) were injected via tail vein into Rag-2 mice that were pretreated with vehicle alone or 110 $\mu\text{g}/\text{day}$ of tamoxifen for 7 days as follows.

25

Rag-2 mice were injected with either PBS (mock) or 4×10^{11} adenovirus particles (as determined by OD_{260}) of type A virus, AD102.7, via the tail vein. At various times after injection, animals were sacrificed and the liver tissues were removed and frozen rapidly on dry ice. To visualize GFP expression in animal tissues, mice were anaesthetized and perfused with 4% paraformaldehyde containing 0.2% glutaraldehyde intracardially (Kafri et al., *Natl. Genet.* 17:314-317, 1997), and the liver tissues were 30 removed and fixed overnight at room temperature in the perfusion buffer containing 30% sucrose. The fixed tissues were sectioned serially and observed under confocal scanning laser microscopy. In experiments evaluating the responses of ligand-regulated recombinase, mice were injected either with vehicle (vegetable oil) alone or with 110 $\mu\text{g}/\text{day}$ of tamoxifen for 7 days prior to adenoviral injection.

Liver tissues from these animals were harvested at 2.5 hrs post injection (the earliest time point taken after injection) and Hirt DNA from approximately 250 mg of frozen hepatic tissue was prepared and analyzed by blot analysis. As shown in Figure 16, the majority of adenoviral DNA was found in circular form in tissues from untreated mice, as well as tamoxifen-treated mice. It can be concluded from these data that the Cre enzyme activity present in the tissue, even in the absence of ligand was sufficient for efficient self rearrangement of virus. As expected, the hepatic tissues from the Rag2 mice injected with AD102.7 showed strong expression of GFP (Figure 17).

Demonstrating that AD102.7 virus, produced efficiently in 293 cells at high titres by the conventional means, can self rearrange efficiently *in vivo* provides the proof of the concept that potentially safer adenoviral gene therapy vectors can be produced.

Adenoviruses Carrying Both FRT and FLP Recombinase

Type A and type B proviral constructs carrying both FRT (FLP recombinase recognition site) and FLP recombinase were also generated. Structures of these viruses are analogous to those of loxP/Cre carrying viruses except that loxP sites are replaced by FRT sites and Cre coding sequence is replaced by FLP coding sequence (Figures 18A, 18B).

Virus Production at Reduced Temperature

Temperature dependence of the E $f1\alpha$ promoter using GFP expression as a marker was also examined. As shown in Table 7, E $f1\alpha$ promoter activity is strongly reduced at 32°C in comparison to 37°C or 39°C. The temperature sensitive nature of the E $f1\alpha$ promoter was used to propagate type B adenovirus carrying FLP at 32°C following initial production of the virus by DNA transfection (pk8-ad3302) at 37°C. HepG2 cells infected with these viruses (AD41.4) showed strong GFP expression, but with an approximately 12 hr delay compared to GFP expressing viruses, suggesting that FLP recombinase activity may be impaired at 37°C. To improve the activity of FLP recombinase, viral constructs were created using a thermostable FLP (referred to as “FLPe”) described by Buchholz *et al.* (Nat. Biotechnol. 16:657-662, 1998).

TABLE 7

Effects of Temperature on Ef1 α Promoter Strength as Shown by GFP Intensities

| Tester plasmids | Arbitrary GFP Intensities | | | |
|--------------------------|---------------------------|-------|-------|-------|
| | | 32°C | 37°C | 39°C |
| Ef1 α GFP | 16 hrs | 1475 | 7886 | 11409 |
| | 41 hrs | 6472 | 36699 | 50787 |
| | 86 hrs | 16256 | 53370 | 54424 |
| Ef1 α Cre +ad2204 | 16 hrs | 243 | 1141 | 2132 |
| | 41 hrs | 1094 | 9119 | 9784 |
| | 86 hrs | 695 | 3219 | 8144 |

5 GFP intensities were measured using a Fluorescent reader.

The activities of FLP and FLPe using a FLP substrate plasmid (ad2879, Figure 19A) in 293 cells were compared. As shown in Table 8, FLPe is significantly more active than FLP under these conditions.

TABLE 8

FLPe is Significantly More Active than FLP Recombinase

| Plasmid | | Mean GFP intensity |
|---------|-------------------|--------------------|
| ad4821 | Ef1 α FLPe | 2.39 |
| ad2949 | Ef1 α FLP | 0.01 |

15 Plasmid coding either FLPe or FLP was cotransfected with a FLP substrate plasmid (Figure 19A) into 293 cells. GFP intensity of each transfection was measured using IP lab program.

20 In addition, a tamoxifen-regulated FLPe was created by fusing the ligand-binding domain from a mutant form of estrogen receptor to the FLPe coding sequence at its C-terminus (FLPe-LBD). The FLPe-LBD was found to be regulated by the ligand, tamoxifen (Table 9). Although FLP activity was retained by C-terminal fusion (FLP-LBD), addition of a short oligopeptide tag to the N-terminus of FLP abolished its activity (data not shown).

TABLE 9

Tamoxifen Regulation of FLPe as Determined by GFP Intensities

| Plasmid | FLPe | -tam | + tam |
|----------------|---------------|------|-------|
| ad4821 +ad2879 | FLPe | ++++ | ++++ |
| ad5022 +ad2879 | FLPe-LBD(tam) | + | +++ |

5

GFP intensity resulting from FLPe mediated recombination was measured using a fluorescent microscope. tam, 2 µg/ml tamoxifen.

10 Inhibition of FLPe Activities by Anti-sense FLPe

An anti-sense approach to inhibit FLP enzyme activity was also employed. This approach tested the notion that incorporation of an open reading frame into an antisense transcript would stabilize the transcript and potentiate antisense activity. Two approaches were utilized. In one approach, the BFP coding sequence was placed upstream of anti-FLPe. In the second approach, an anti-FLPe was placed upstream of an internal ribosome entry sequence (IRES) and the BFP coding sequence (Figure 20). The ability of these constructs to inhibit FLPe was assayed using a FLP substrate plasmid, ad2879 (Figure 19A) and the result is shown in Figure 21. These data show that anti-sense FLPe is more effective in inhibiting FLPe function when it is fused to BFP, which can presumably be replaced with any other stable protein.

OTHER SELF-REARRANGING ADENOVIRUSESMixed Infection With Adenoviruses Carrying loxP/FLP and FRT/Cre

25 One of the ways to produce adenoviruses that can be rearranged in target cells but not producer cells is to engineer two separate viruses, each carrying one recombinase and the target sequence for the other. To test this system, type B adenoviral constructs carrying Cre recombinase and FRT sites and FLPe recombinase and loxP sites were created (Figure 22). In target cells infected with both viruses, Cre catalyzes 30 recombination between the two loxP sites in the FLP virus, and FLP carries out FRT mediated recombination in the Cre virus, resulting in two circular plasmids. The loxP virus contained BFP, whereas the FRT virus contained GFP. Measurement of the fluorescent intensities of GFP and BFP, after cotransfected the two constructs, revealed that BFP expression (mediated by Cre enzyme) was greater than GFP expression

(mediated by FLP enzyme), suggesting that the Cre enzyme functions more efficiently than FLP.

Accordingly, these two recombinase activities in the target cells need to be balanced for complete circularization of both viral vectors. Exemplary methods for 5 modulating Cre /FLP activity include the use of transcriptional regulation (such as by varying promoter strength and/or with or without poly A addition signal sequence) and translational and/or post-translational regulation (such as by changing FLP to FLPe and making LBD fusion proteins), and post viral production control (such as by changing the ratio of two viruses).

10 In one approach, Cre was replaced by Cre-LBD and FLP was replaced by FLPe. To improve identification of the rearrangement products, BFP was replaced with RFP as a marker for Cre recombination. As shown in Table 10, in the presence of estrogen, expression of GFP (FLPe mediated) and RFP (Cre mediated) were similar.

15

TABLE 10

RFP and GFP Expression of Cells Cotransfected With Type B Proviral Constructs Carrying Cre-LBD/FRT(GFP) or FLPe/loxP (RFP)

20

| Plasmids | Genotypes | GFP intensity | | RFP intensity | |
|---------------------------|--|---------------|------|---------------|------|
| | | Estrogen | | | |
| pk8-ads120 +pk8-ads113 | Cre-LBD/FRT (GFP) + FLPe/loxP (RFP) | - | + | - | + |
| | | 3222 | 3183 | 46 | 1954 |

To insure that both Cre and FLP carrying viruses with an optimal ratio infect each target cell, these viruses can be cross-linked prior to infection. For example, Cre carrying virus is labeled by biotin while FLP carrying virus is labeled by avidin. Mixing 25 two types of modified viruses generates virus complexes of desired proportions as well. Biotinylation or avidinylation can be carried out using commercially available reagents such as EZ-Link TFP-PEO biotin (Pierce) and EZ-Link maleimide activated NeutrAvidin (Pierce). The extent of the biotin/virus and avidin/virus will be empirically determined to ensure the viability of the virus and to obtain an optimal ratio of two 30 viruses in the complex. Optimal ratios will be those resulting in 1:1 Cre and FLP recombinase activities in target cells. The modifications will be done following manufacturer's instructions.

This approach not only increases the effective capacity of adenoviral vector but also opens new avenue of applications involving multiple proteins, some of which cannot be coexpressed in production cell line as a result of combination toxicity.

- 5 All references mentioned herein are hereby incorporated by reference.
Other embodiments are within the claims.

What is claimed is:

10

Claims

- 5 1. A replicatable viral DNA vector encoding a site-specific DNA-altering enzyme and a DNA target recognized by said enzyme, said enzyme selectively converting, in a cell expressing said enzyme, said DNA vector to a rearranged form.
- 10 2. The vector of claim 1, wherein said rearranged form comprises an autonomously replicating episome.
- 15 3. The vector of claim 1, wherein said rearranged form comprises linear and circular DNAs.
- 20 4. The vector of claim 1, wherein said vector comprises adenoviral DNA.
- 25 5. The vector of claim 1, wherein said vector comprises a genetically-engineered recombination site.
- 30 6. The vector of claim 5, wherein said recombination site comprises a target of Cre or FLP.
7. The vector of claim 1, wherein said enzyme comprises a recombinase or an integrase.
8. The vector of claim 7, wherein said recombinase is Cre or FLP recombinase.
9. The vector of claim 1, wherein said enzyme is functional in a mammalian cell.
10. The vector of claim 5, wherein said recombination site comprises a recognition sequence of a site-specific DNA-altering enzyme.

11. The vector of claim 1, wherein said vector comprises an origin of replication functioning in a mammalian cell.
12. The vector of claim 11, wherein said origin of replication is an Epstein Barr Virus replicon.
5
13. The vector of claim 1, wherein said vector comprises a gene of interest.
14. A method for assembling a recombinant adenoviral DNA said method comprising the steps of: (a) providing a first linearized DNA vector comprising a restriction site and a cos site and a second linearized DNA vector comprising said restriction site, an adenoviral nucleic acid molecule, and a cos site; and (b) ligating said first and second linearized DNA vectors, said ligation assembling a recombinant adenoviral DNA.
10
15. The method of 14, wherein said first linearized DNA vector comprises a selectable marker.
15
16. The method of claim 14, wherein said first linearized DNA vector comprises an adenoviral left end-inverted terminal repeat.
20
17. The method of claim 14, wherein said first linearized DNA vector comprises a gene of interest.
18. The method of claim 14, wherein said second linearized DNA vector comprises a selectable marker.
25
19. The method of claim 14, wherein said second linearized DNA vector comprises an adenoviral right-end inverted terminal repeat.
20. The method of claim 14, said method further comprising packaging said assembled adenoviral DNA into a phage and infecting a host cell.
30

21. The method of claim 14, wherein said first and second linearized DNAs comprise a cosmid vector.
 22. The method of claim 14, wherein said adenoviral DNA is flanked by 5 cleavage sites.
 23. The method of claim 22, wherein said cleavage sites comprise intron endonuclease cleavage sites.
- 10 24. An adenovirus producer cell comprising a nucleic acid molecule that expresses a dominant negative site-specific DNA-altering enzyme.
25. The producer cell of claim 24, wherein said site-specific DNA altering enzyme is a dominant negative recombinase.
- 15 26. The producer cell of claim 25, wherein said recombinase is a Cre or Flp recombinase.
- 20 27. The producer cell of claim 26, wherein said dominant negative recombinase is CreY324C.
28. The producer cell of claim 26, wherein said Flp recombinase is Flpe.
29. The producer cell of claim 24, wherein said cell is a 293 human embryonic kidney cell.
30. A vector comprising, in the 5' to 3' direction, a first genetically engineered *cis*-acting target recognized by a site-specific DNA altering enzyme;
- 30 a gene of interest;
- 30 a lineage-specific gene promoter;
- 30 a second genetically engineered *cis*- acting target recognized by a site-specific DNA altering enzyme; and
- 35 a nucleic acid molecule encoding a site-specific DNA altering enzyme.

31. A vector comprising, in the 5' to 3' direction,
a first genetically engineered *cis*-acting target recognized by a site-
specific DNA altering enzyme;
a gene of interest;
5 a bi-directional promoter, comprising a second genetically engineered
cis-acting target recognized by a site-specific DNA altering enzyme; and
a nucleic acid molecule encoding a site-specific DNA altering enzyme.

10 32. A method of gene therapy comprising the administration to a patient in
need of gene therapy a therapeutically effective amount of the vector of any one
of claims 1, 30, or 31 which is expressed in said patient

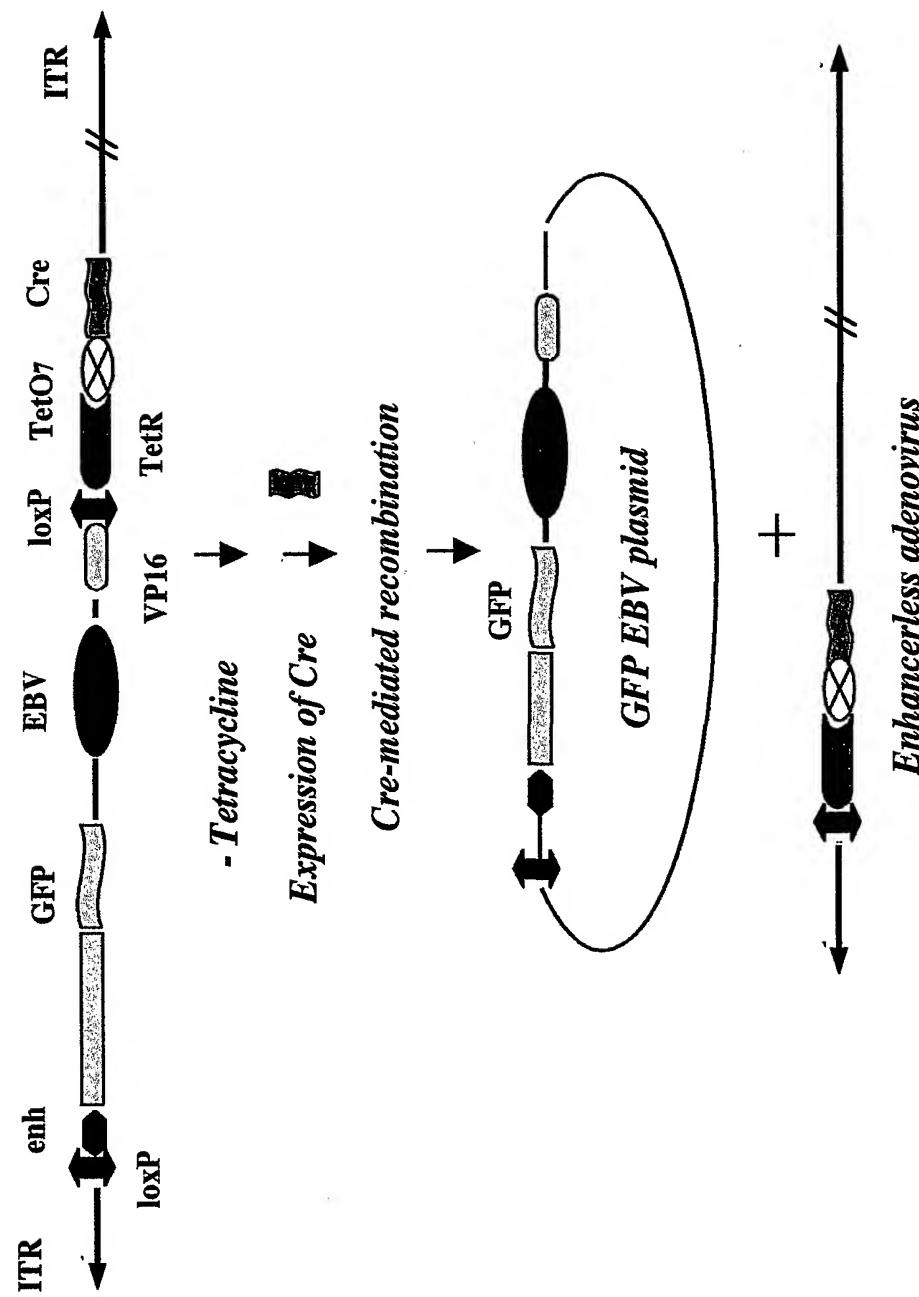
15 33. A population of cells transfected with the vector of any one of claims 1,
30, or 31.

15 34. A method of gene therapy comprising the administration to a patient in
need of gene therapy a therapeutically effective amount of the population of cells
of claim 33.

20

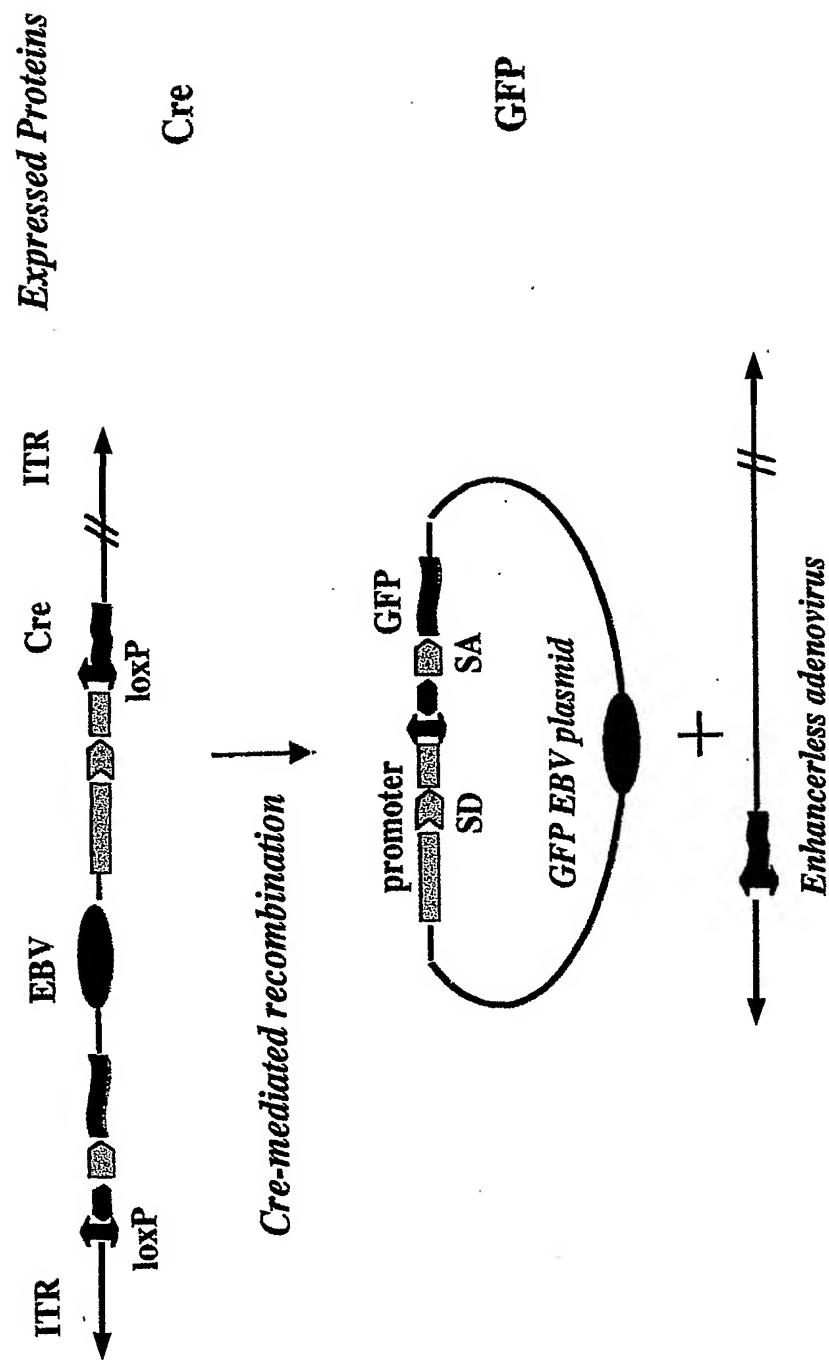
1/40

FIG. 1A



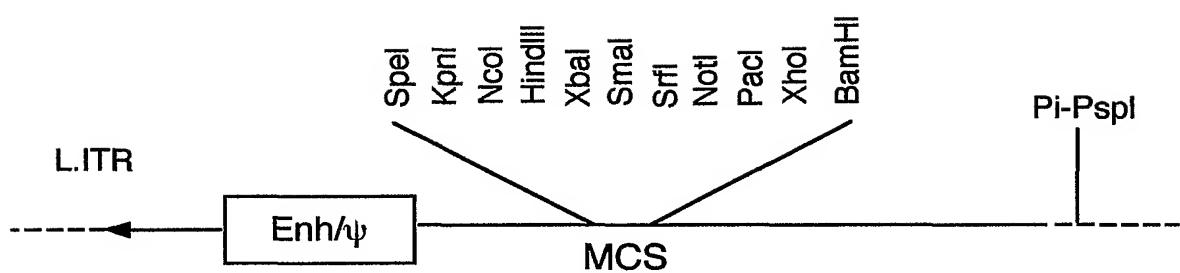
2/40

FIG. 1B



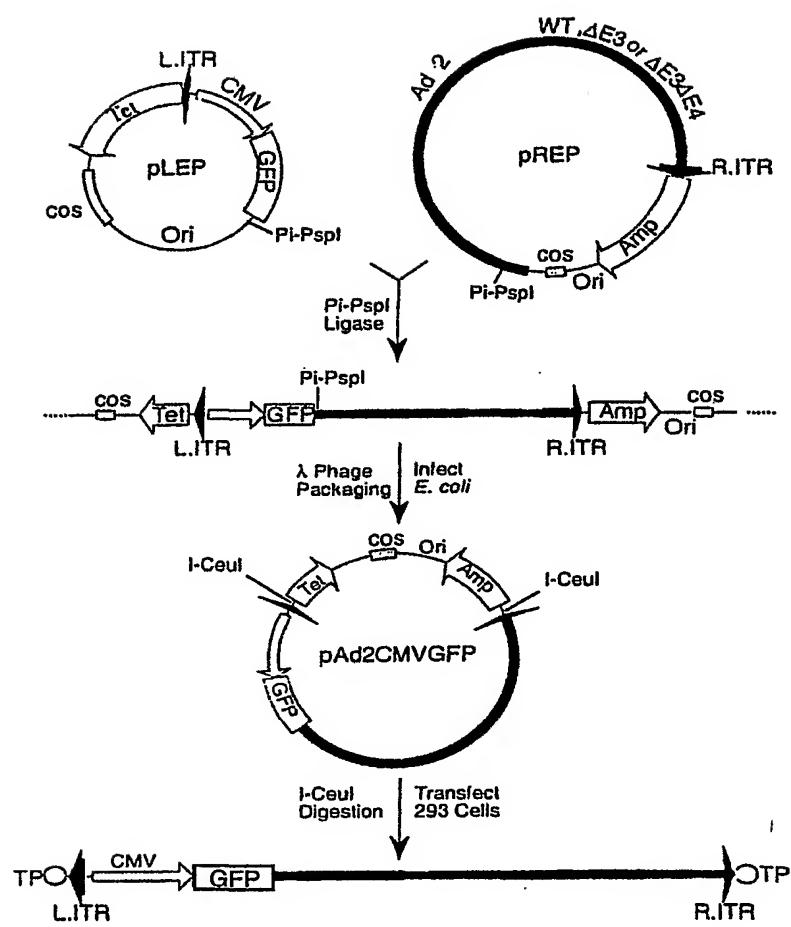
3/40

FIG. 2A



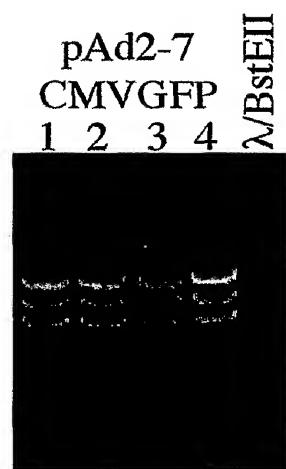
4/40

FIG. 2B



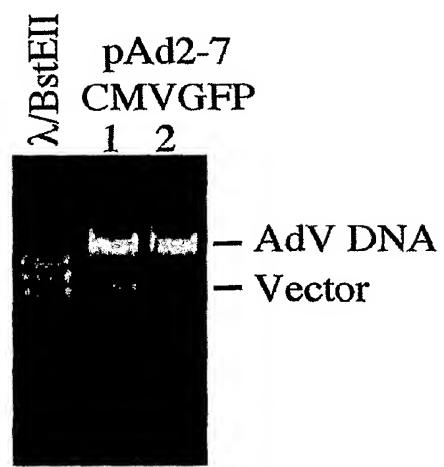
5/40

FIG. 3A



6/40

FIG. 3B



7/40

FIG. 4A

| Ad DNA forms | Digestion | CPEs | Time (days) |
|--|------------------------------|------|----------------|
| 1. Circular Ad DNA | pIAd2B | 0.17 | -- |
| 2. Linearized cosmid DNA | pIAd2B / I-CeuI | 8.5 | 10 |
| 3. Liberated AdV DNA from cosmid | pIAd2B / I-CeuI +BsaBI | 76 | 6 |

8/40

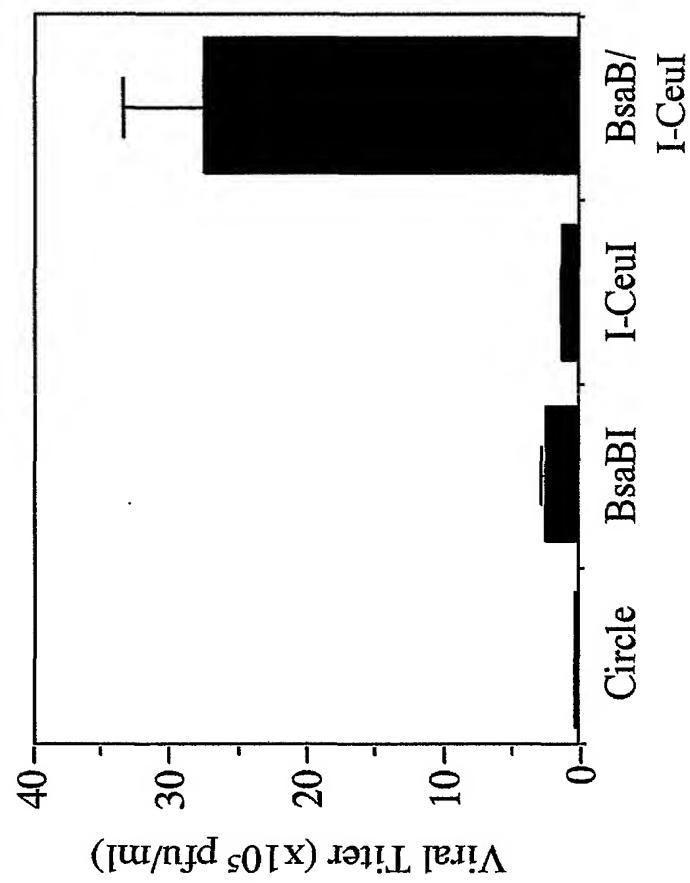
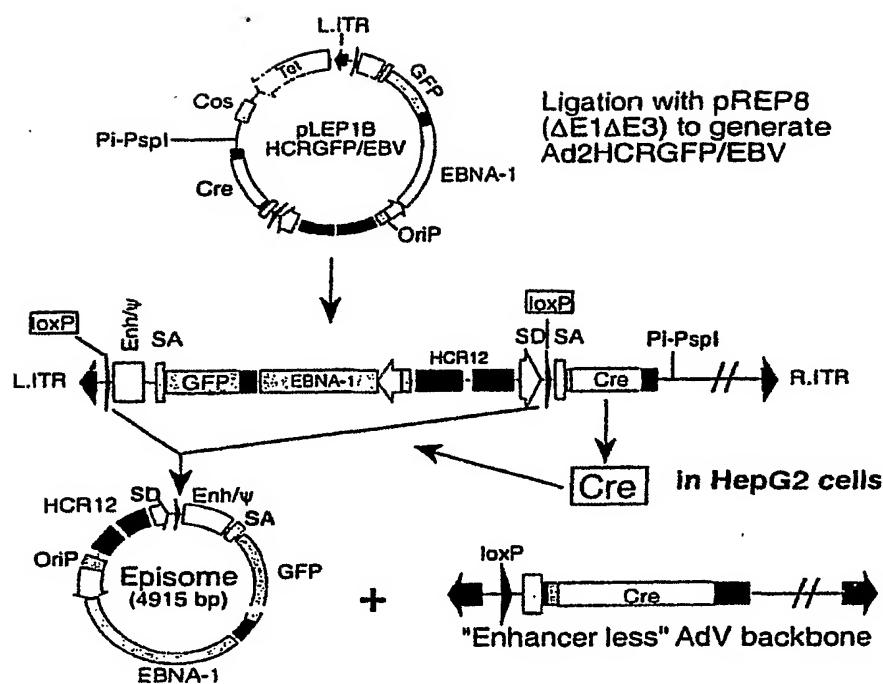


FIG. 4B

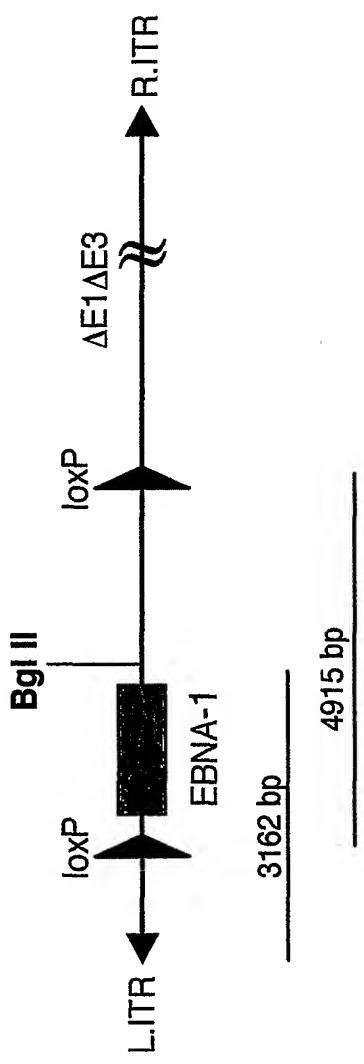
9/40

FIG. 5



10/40

FIG. 6A



11/40

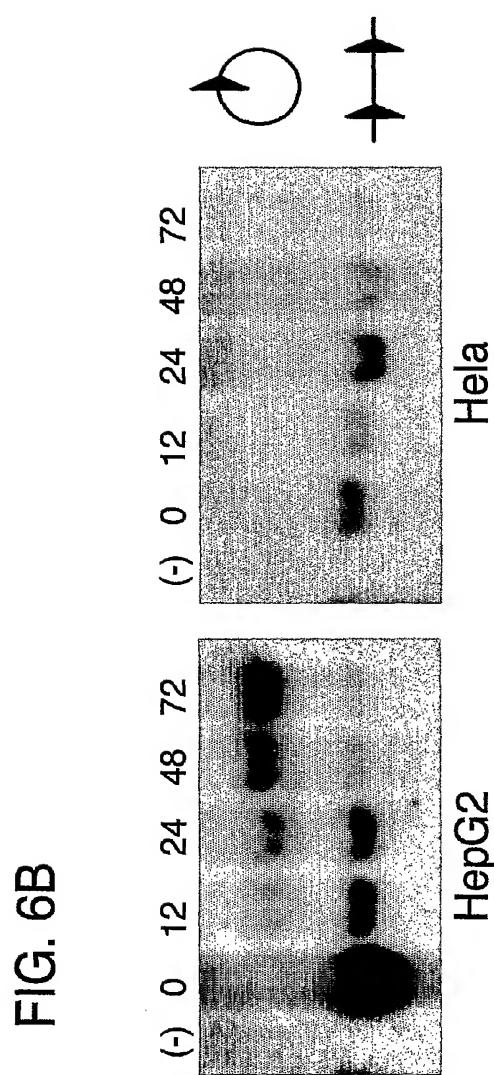
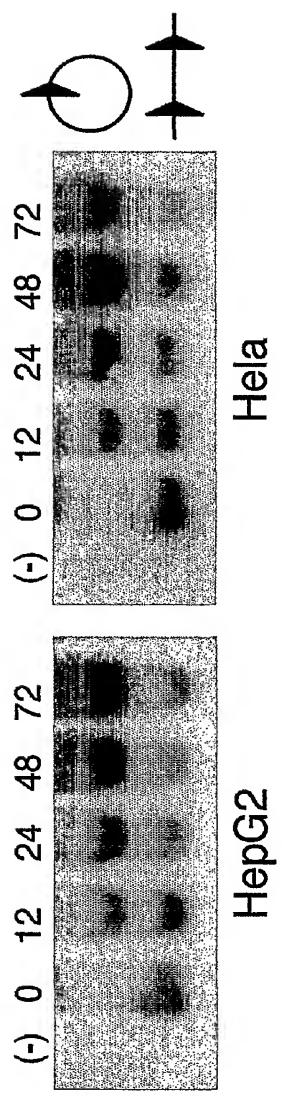


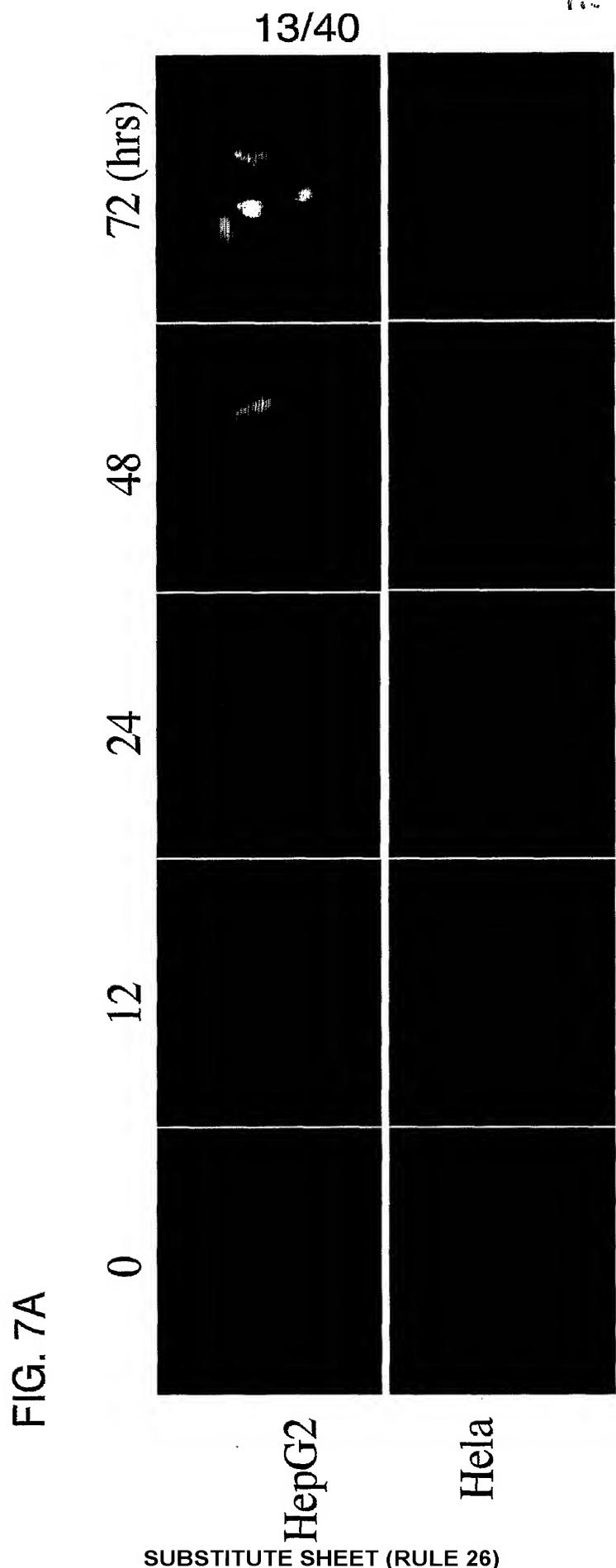
FIG. 6B

12/40

FIG. 6C

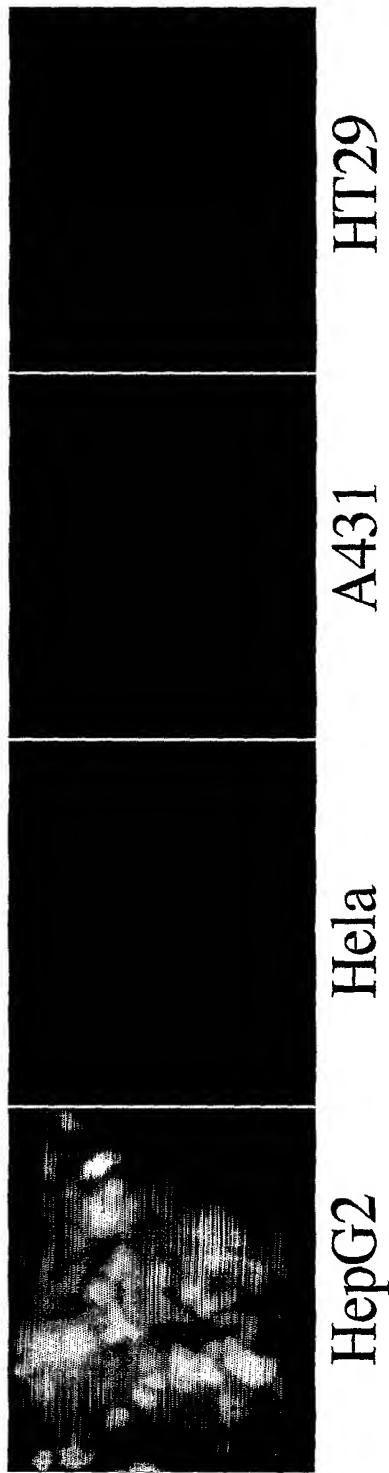


PC 29 NO



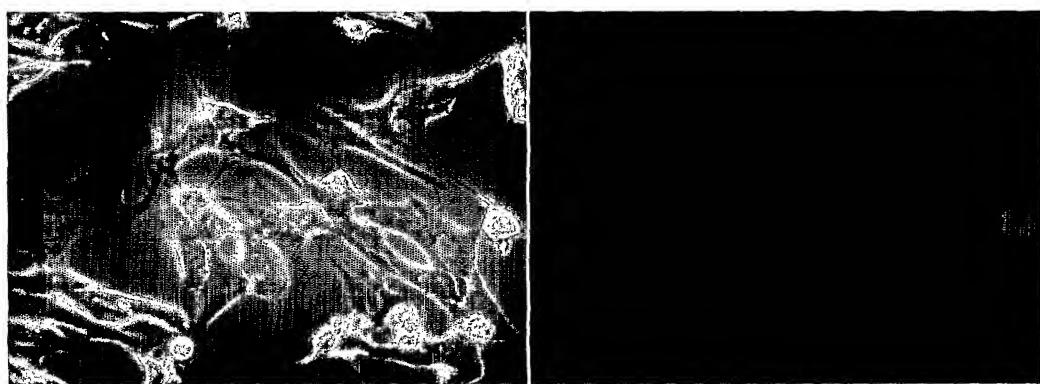
14/40

FIG. 7B



15/40

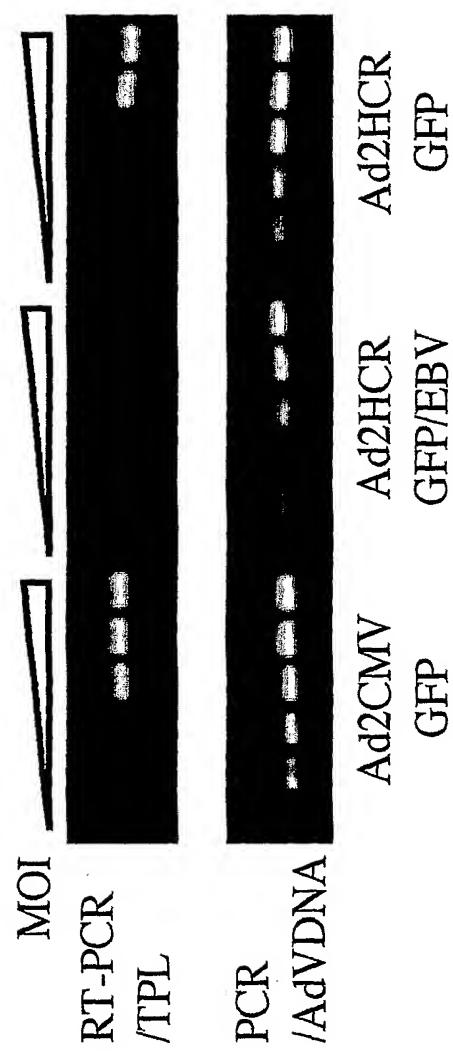
FIG. 7C



Human primary hepatocytes

16/40

FIG. 8A



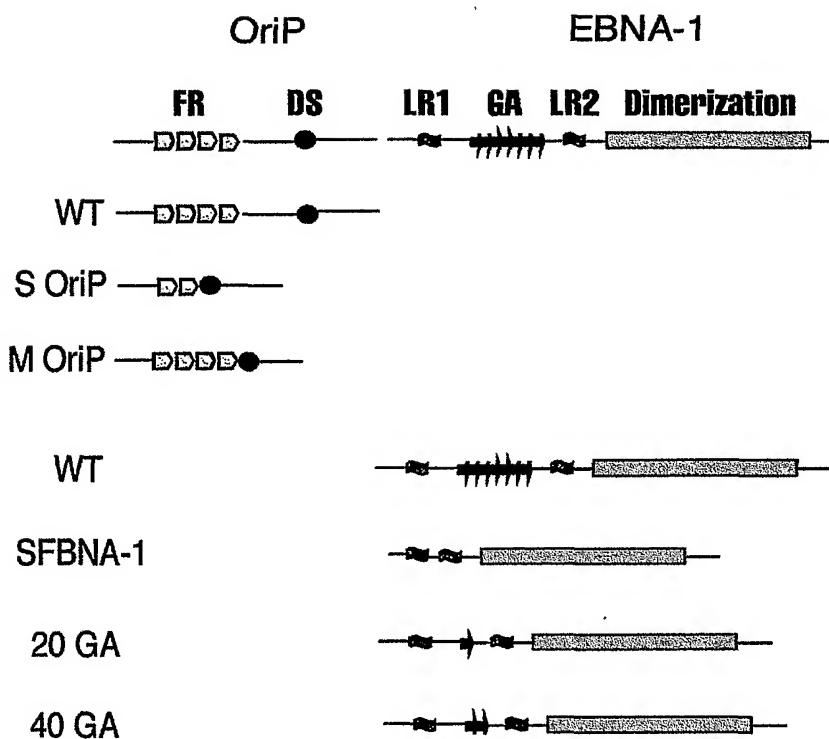
17/40

FIG. 8B

| # of TPL mRNA / 10^6 AdV DNA | | |
|--------------------------------|-----------------------------|--|
| Ad2HCRGFP | $1.15 \pm 0.28 \times 10^4$ | |
| Ad2HCRGFP/EBV | Non-detectable | |

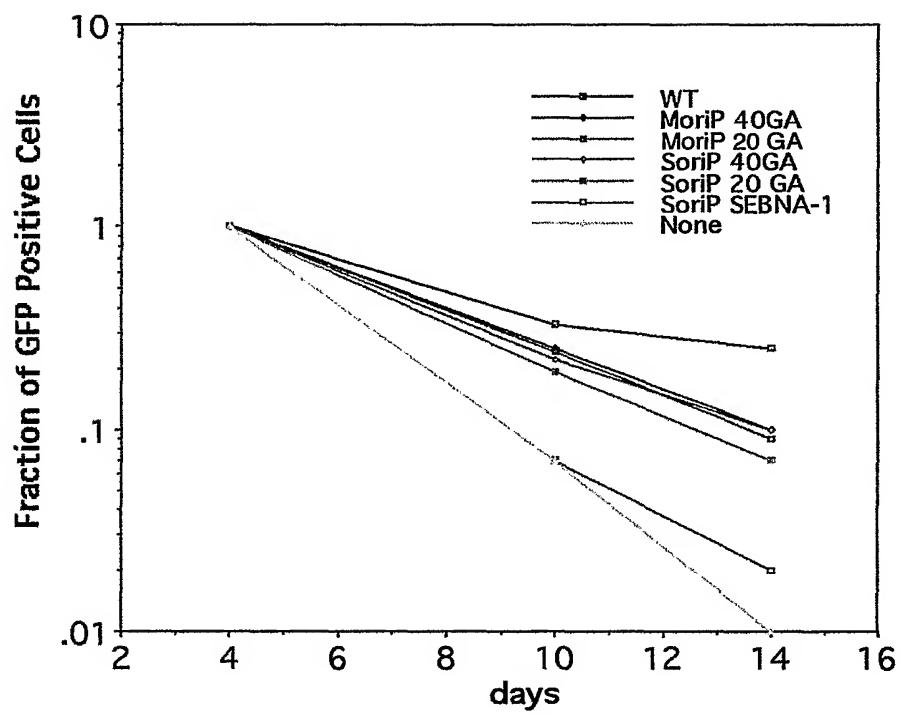
18/40

FIG. 9A



19/40

FIG. 9B



20/40

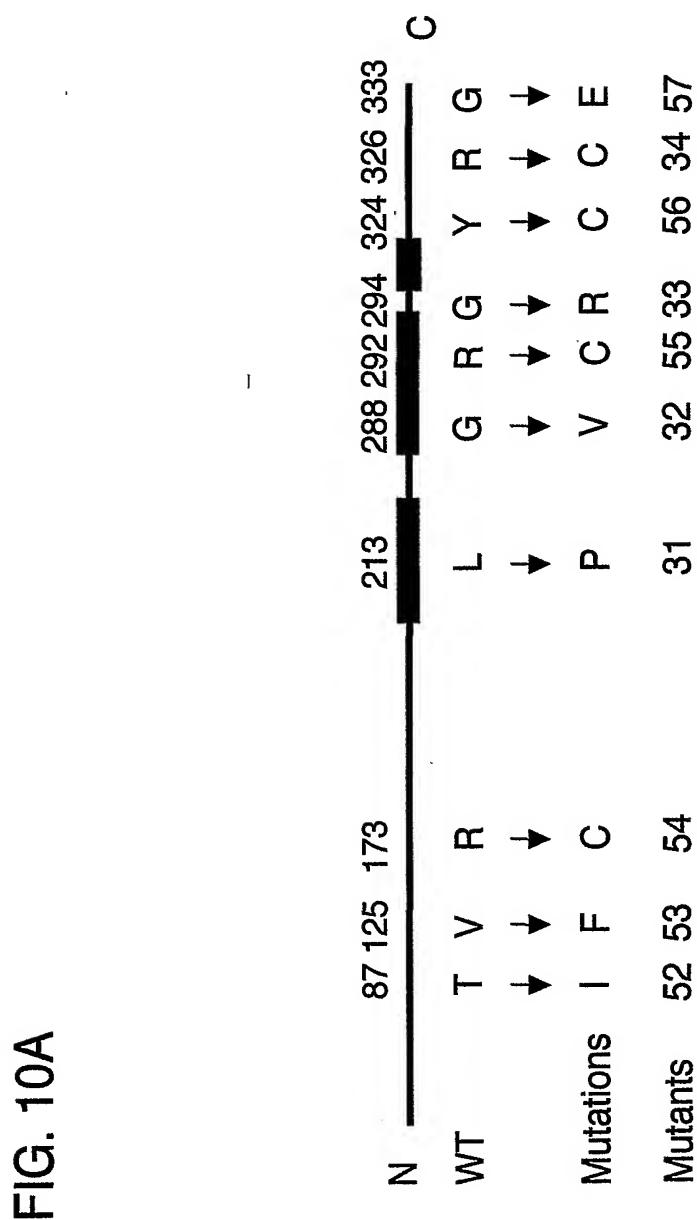
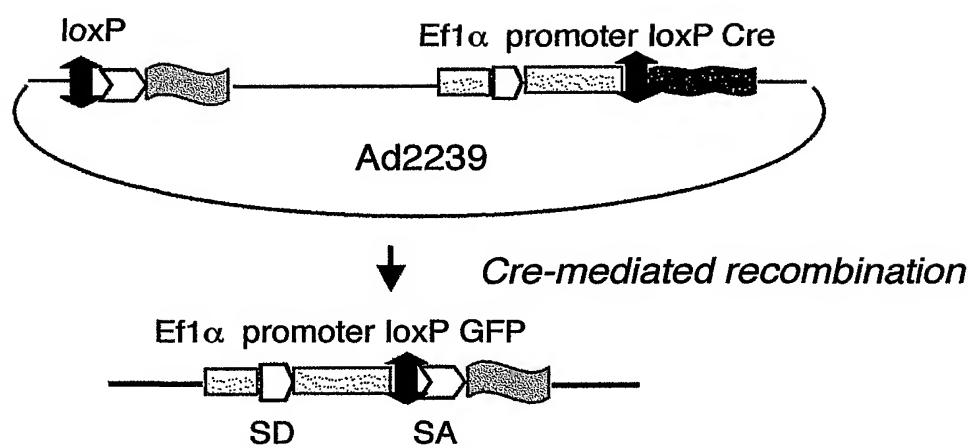


FIG. 10A

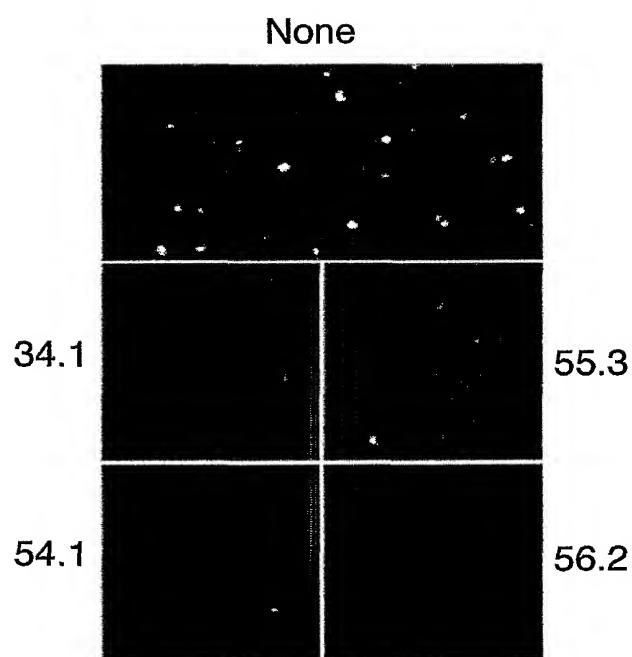
21/40

FIG. 10B



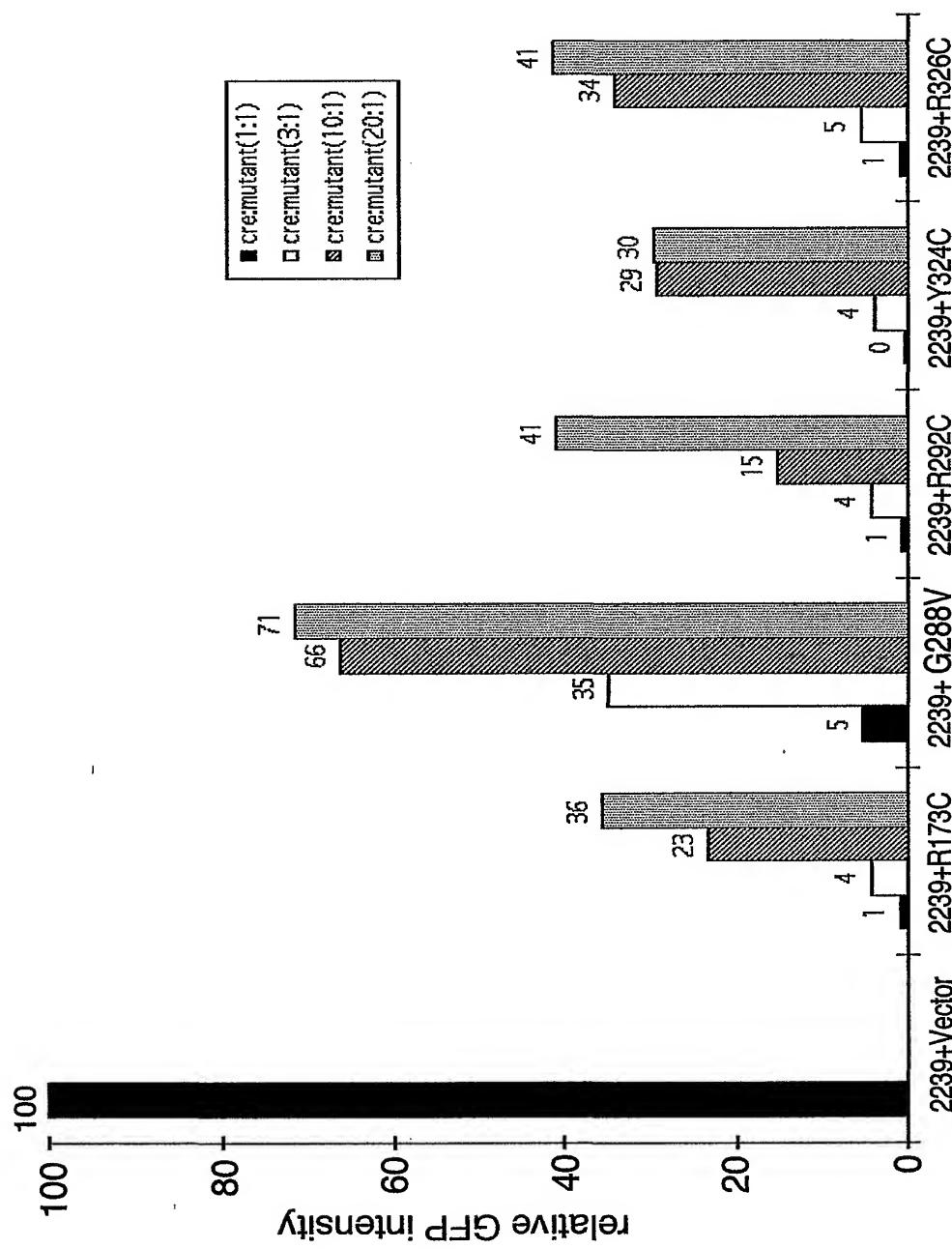
22/40

FIG. 10C



23/40

FIG. 10D



24/40

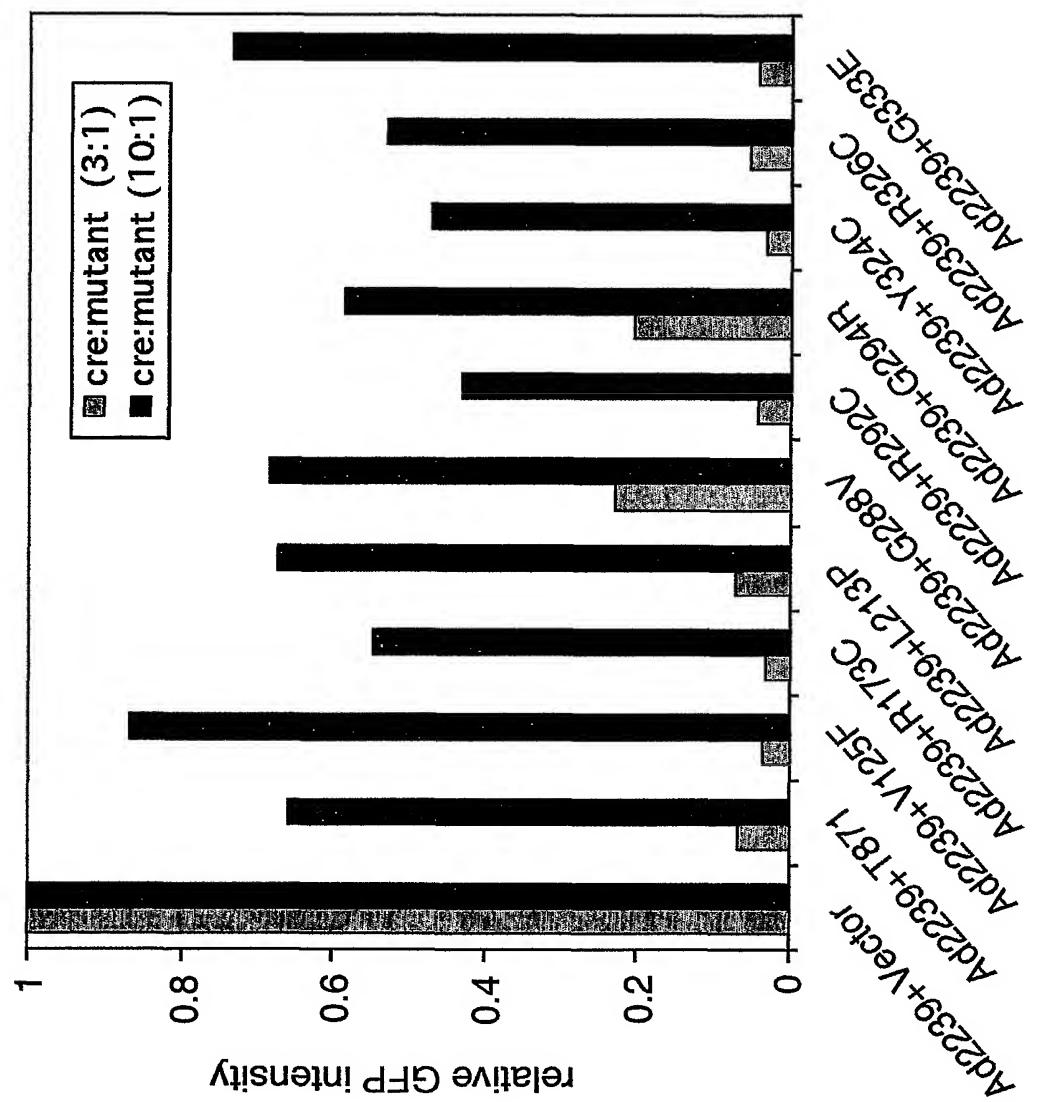
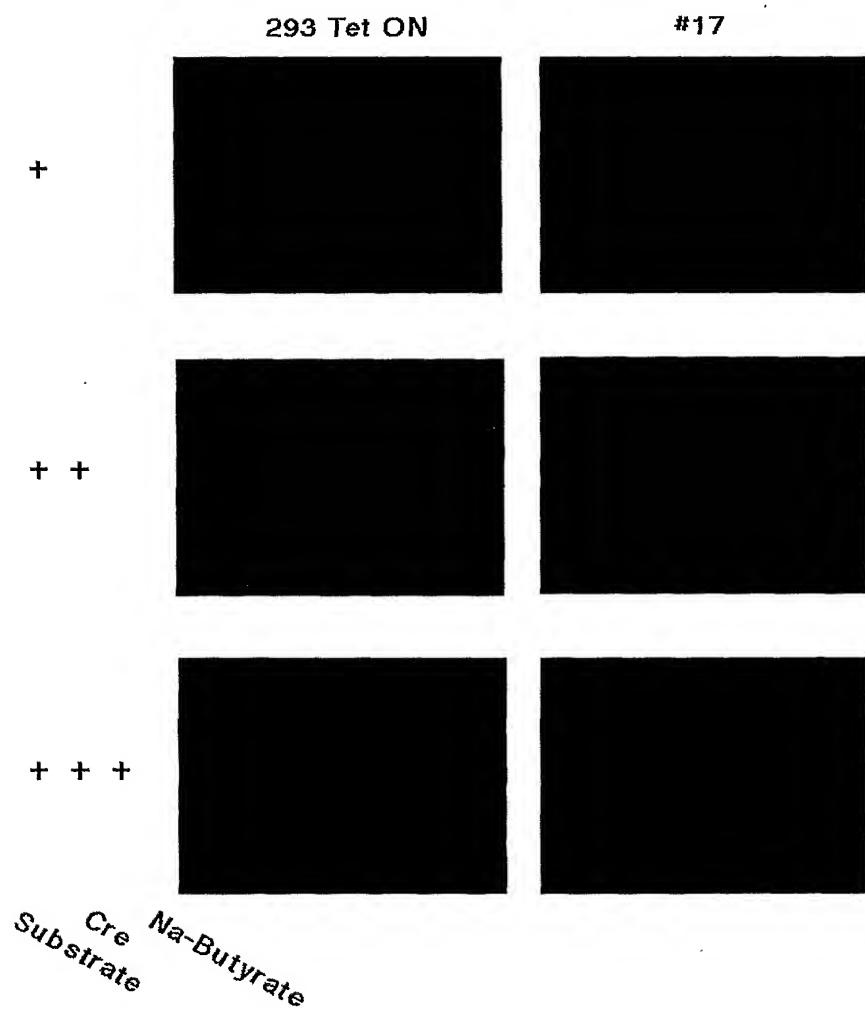


FIG. 10E

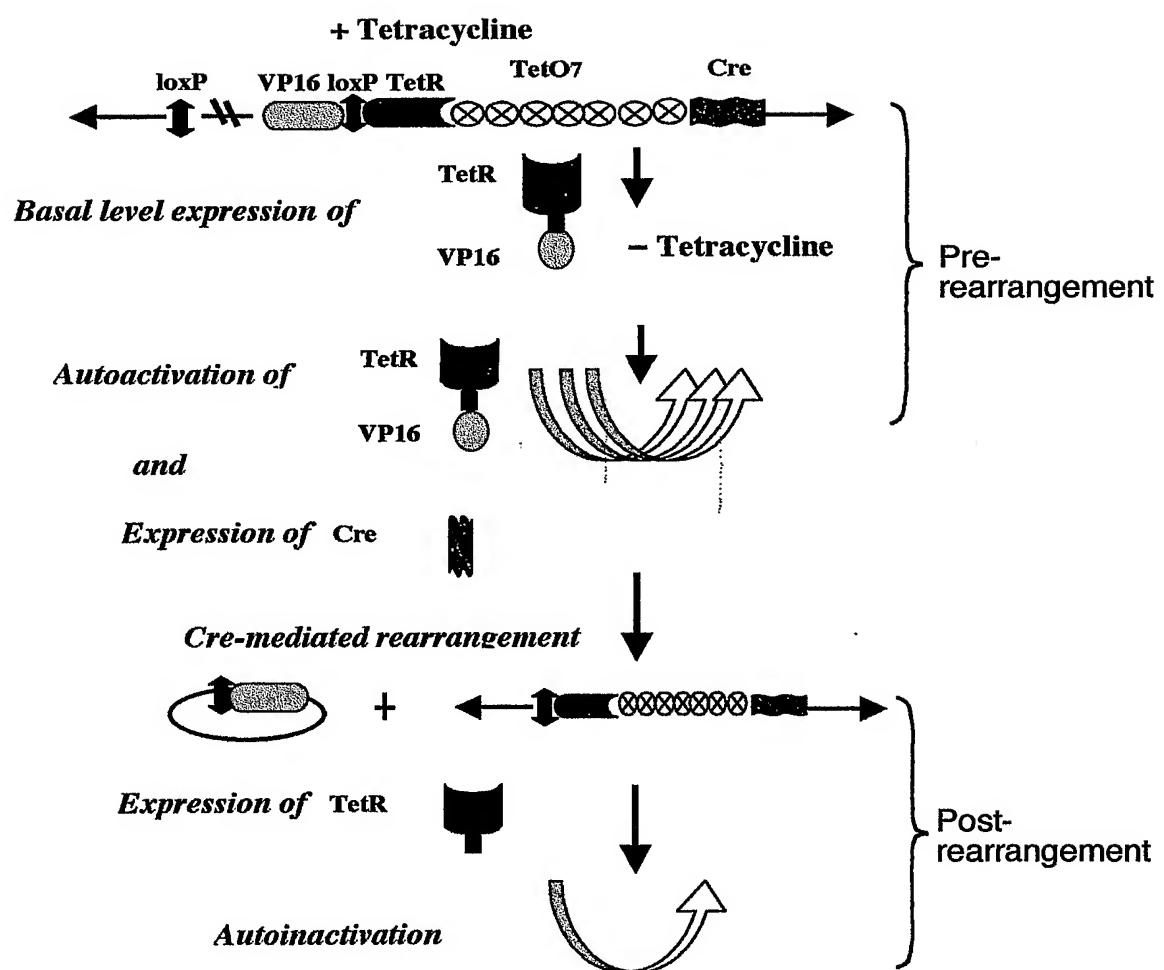
25/40

FIG. 11



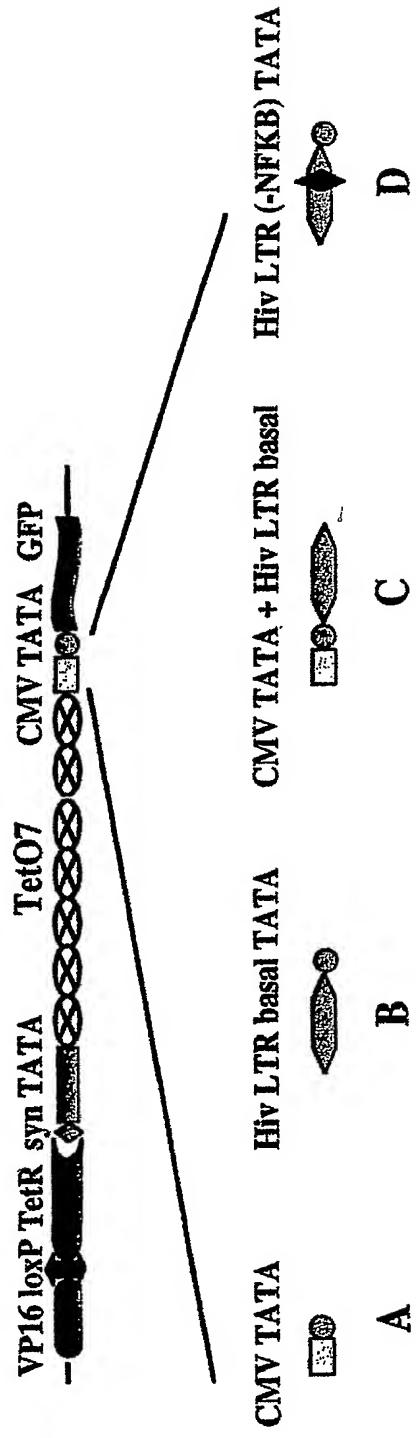
26/40

FIG. 12



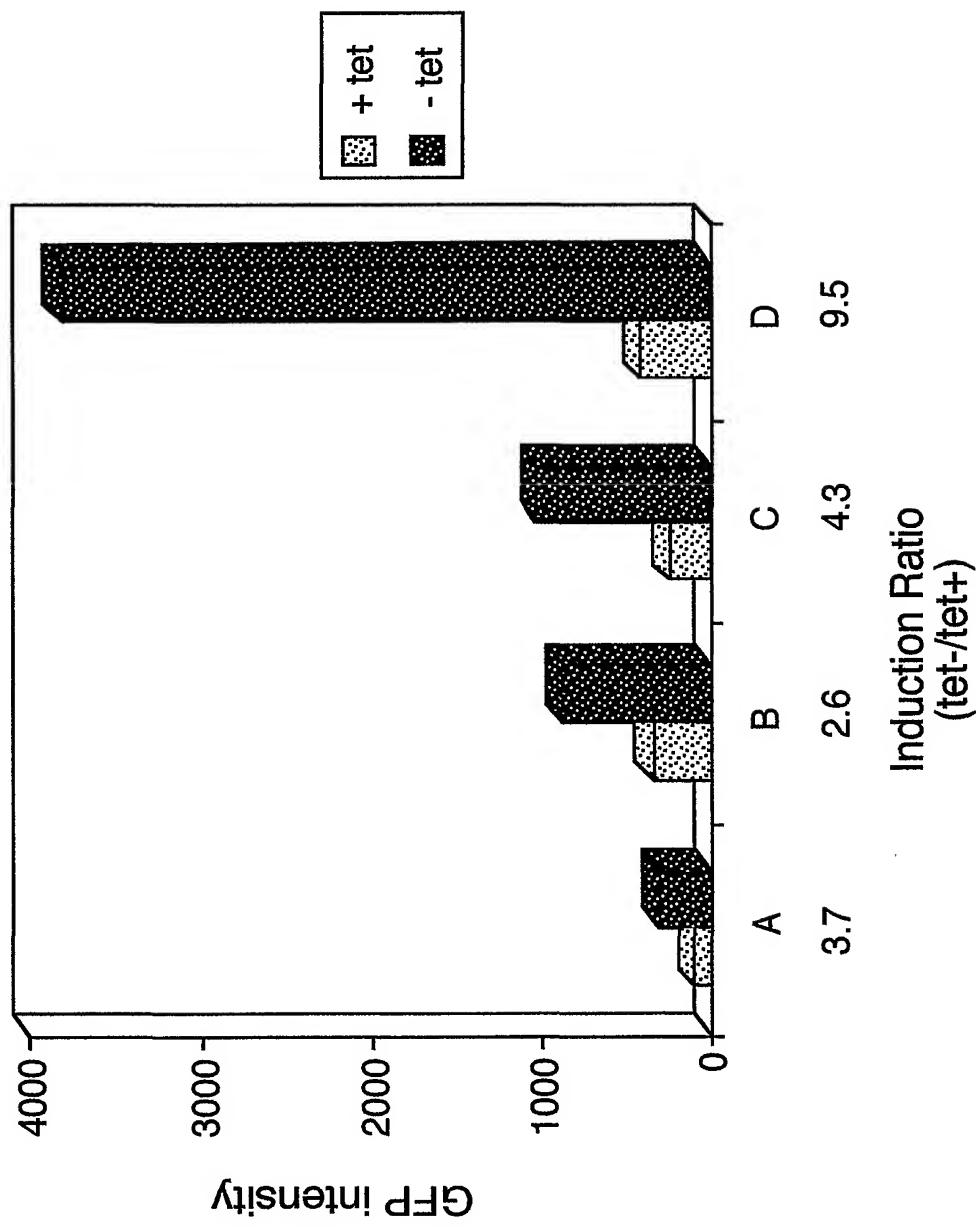
27/40

FIG. 13A



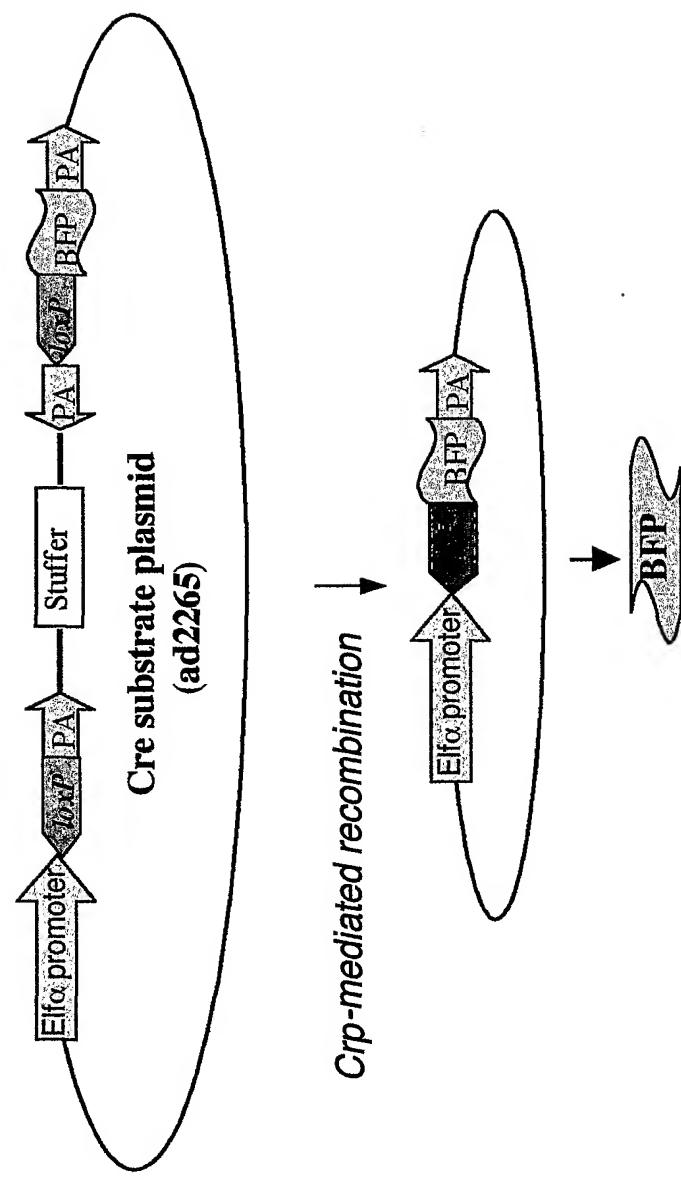
28/40

FIG. 13B



29/40

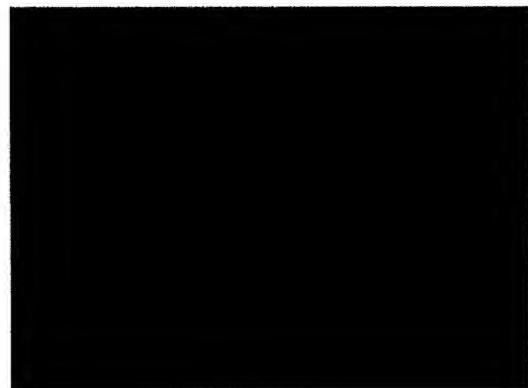
FIG. 14



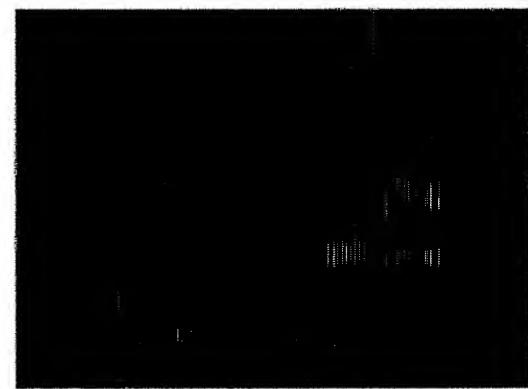
30/40

FIG. 15A

0 μ M Est.

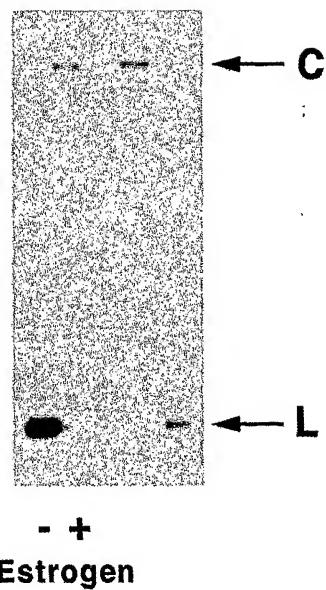


2 μ M Est.



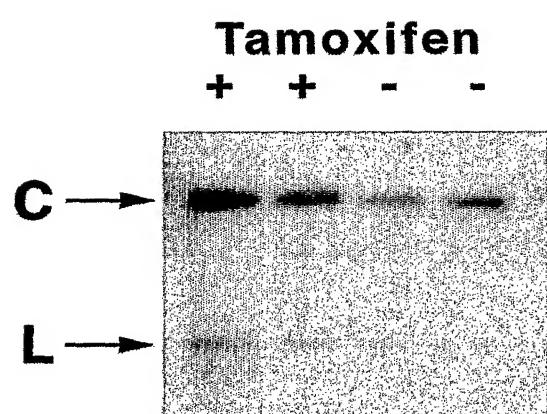
31/40

FIG. 15B



32/40

FIG. 16



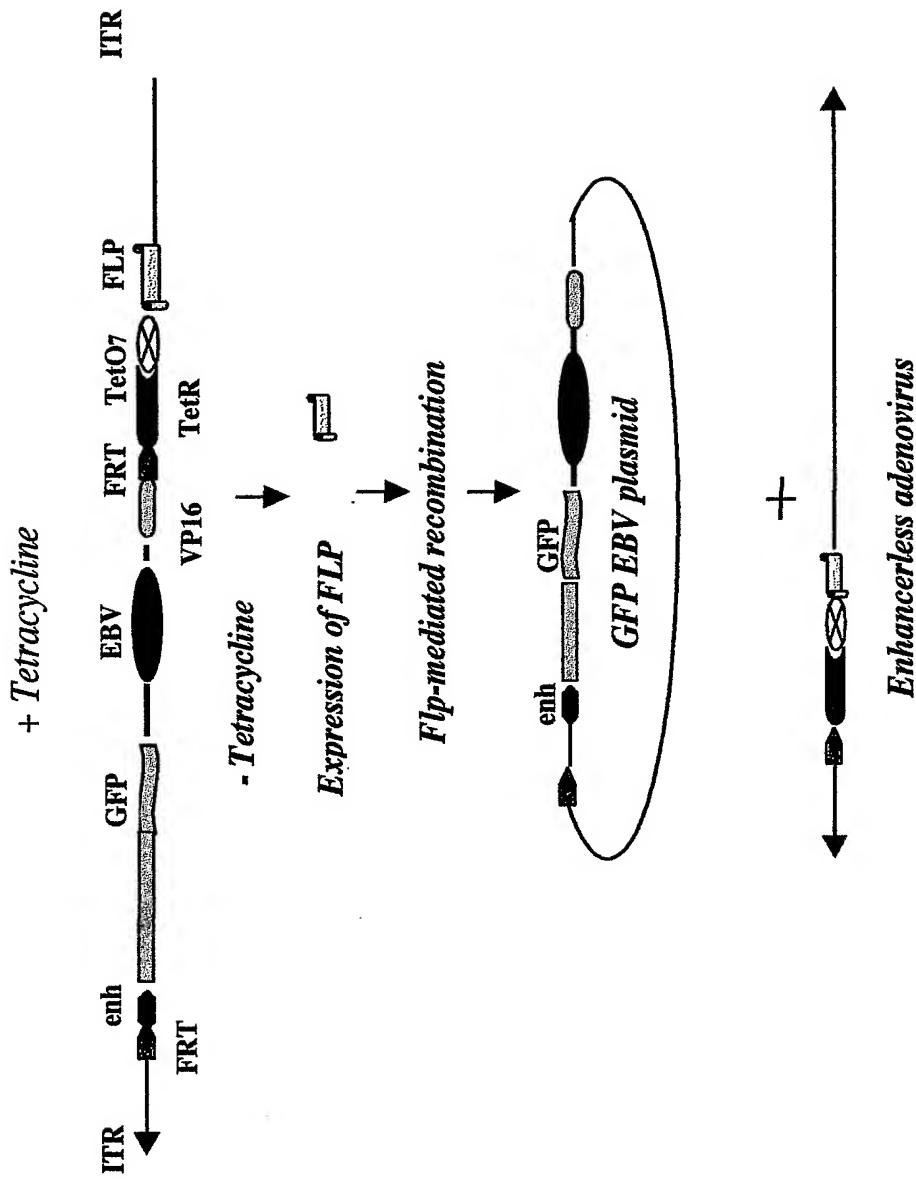
33/40

FIG. 17



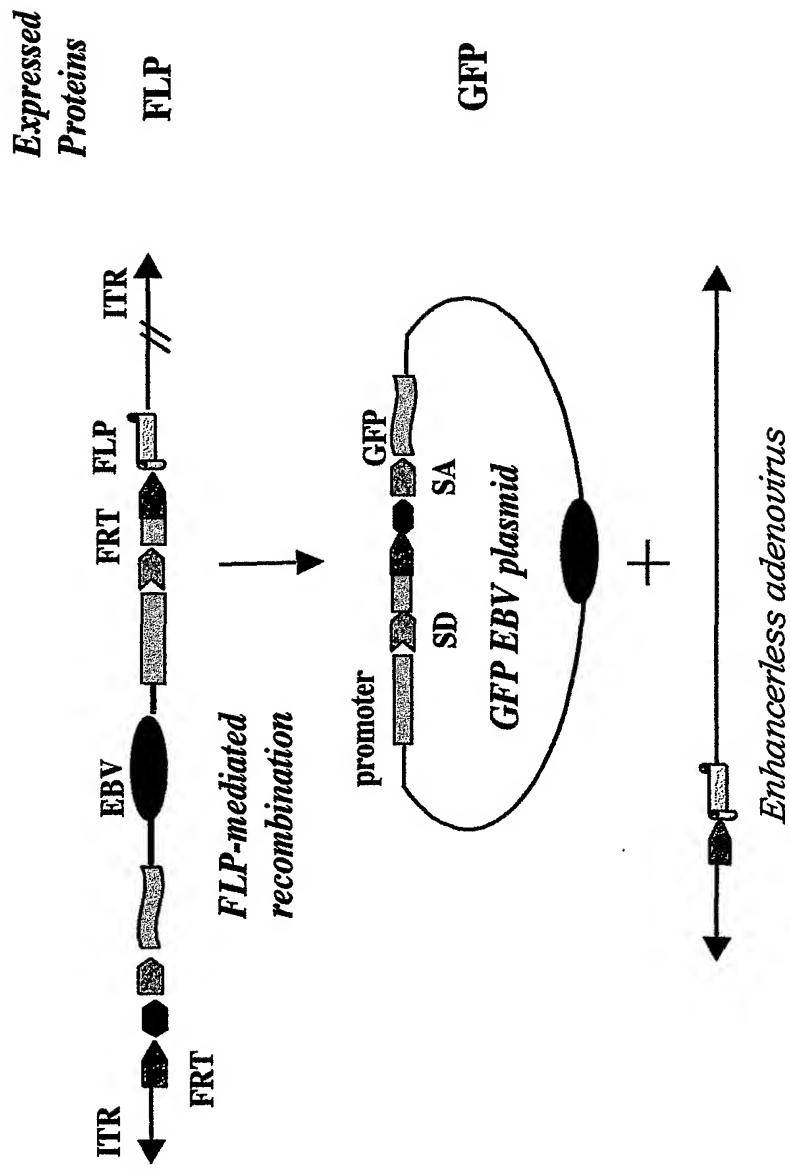
34/40

FIG. 18A

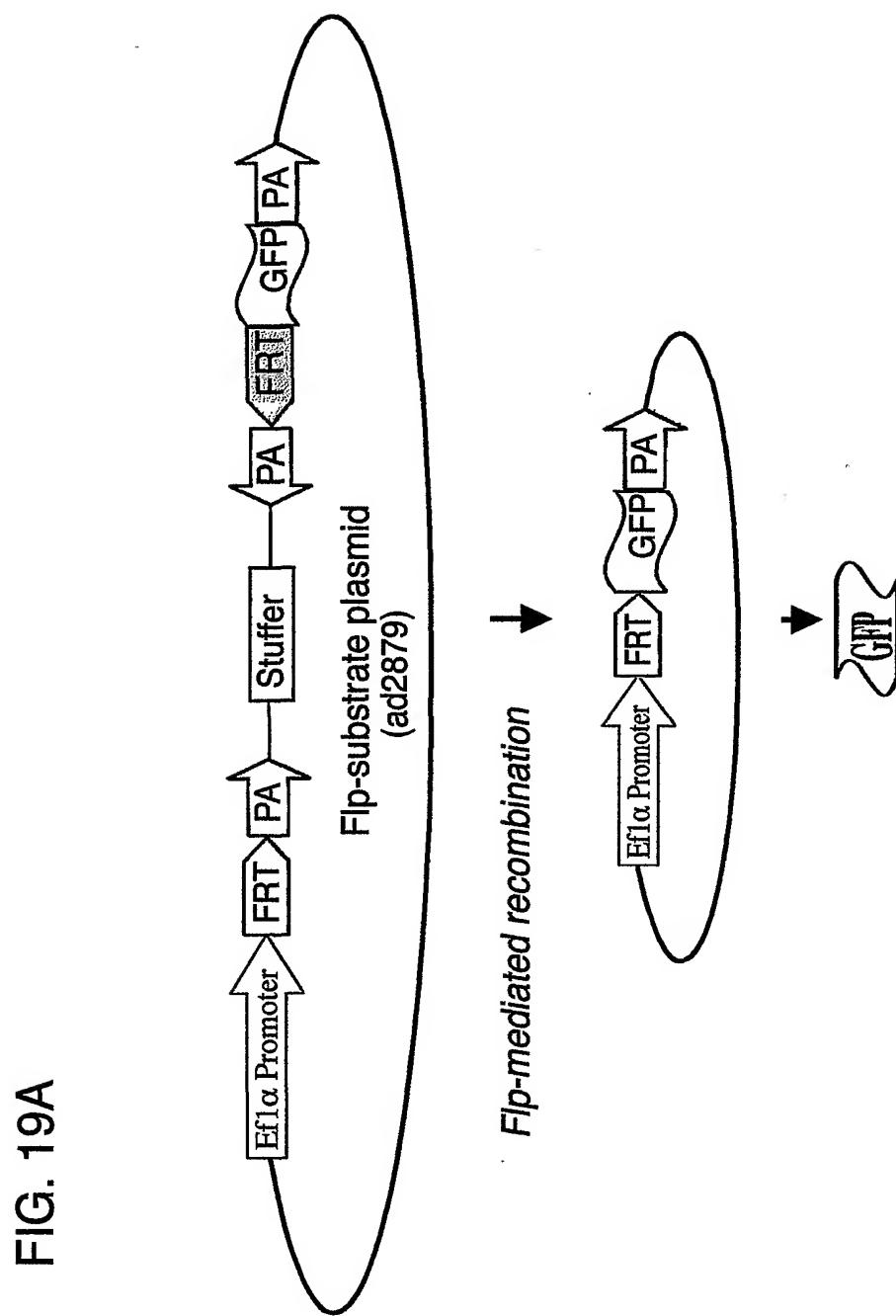


35/40

FIG. 18B

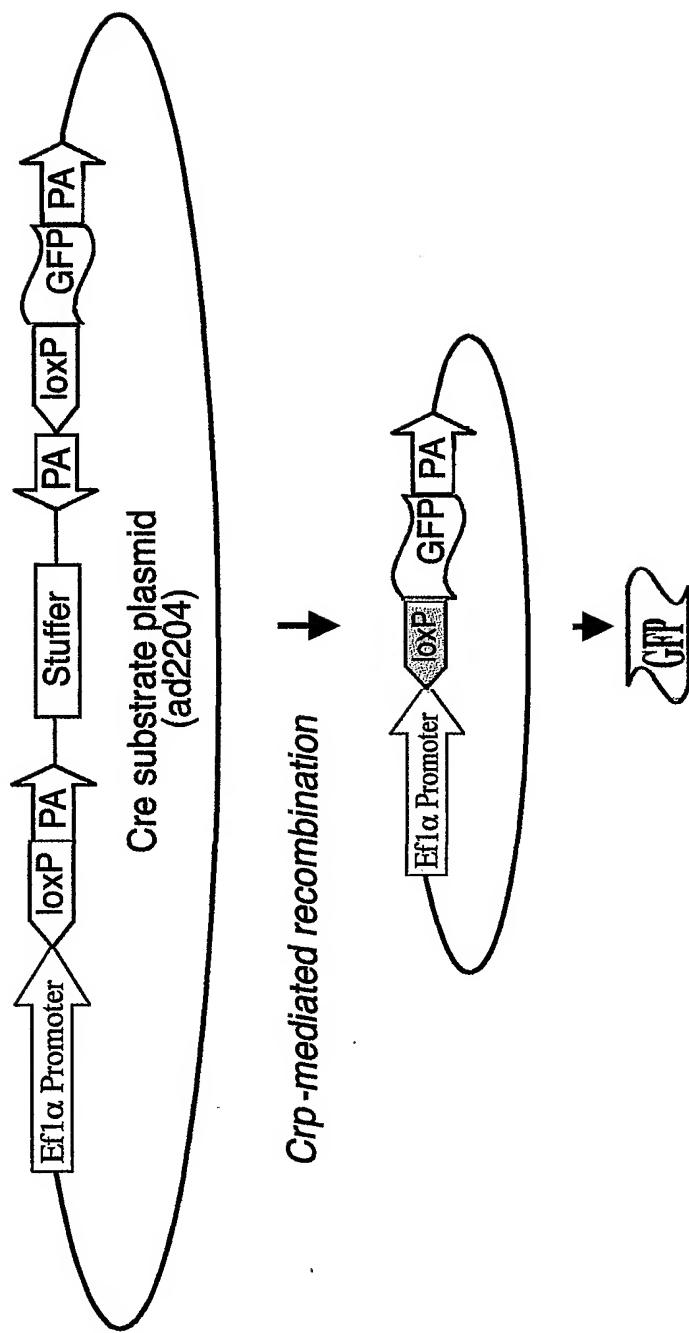


36/40



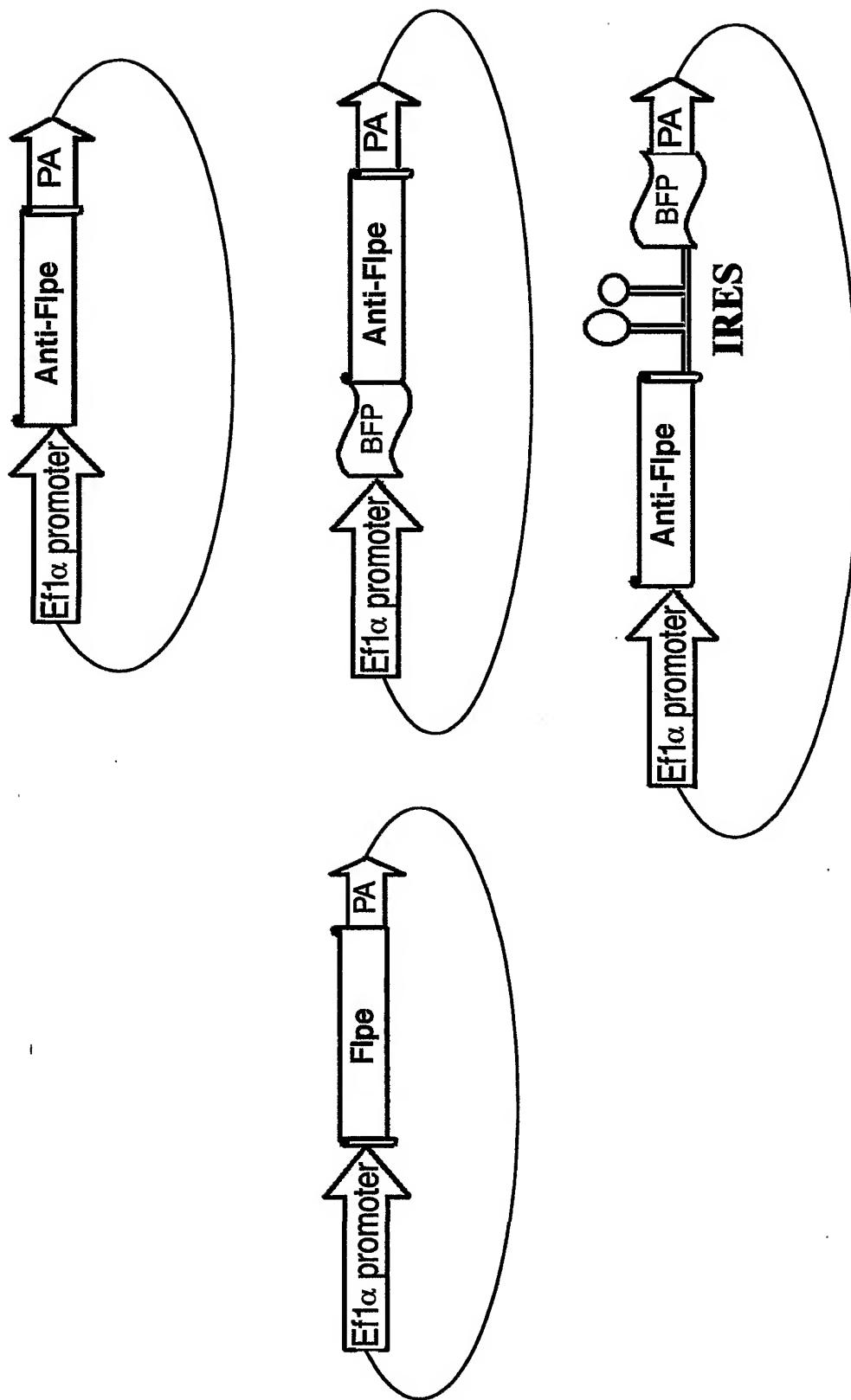
37/40

FIG. 19B



38/40

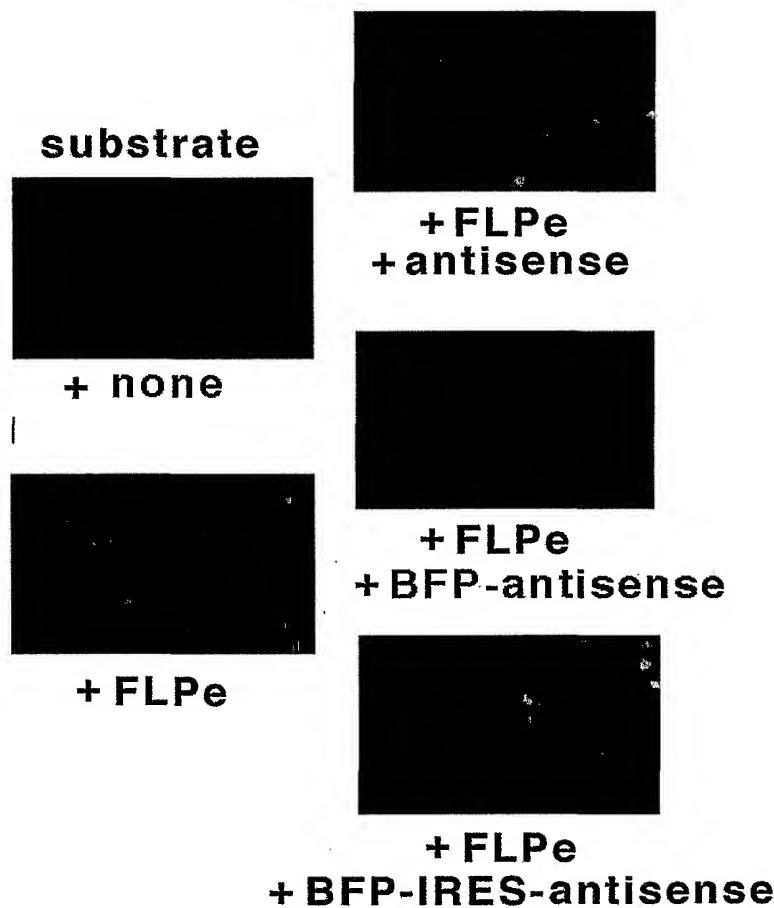
FIG. 20



M

39/40

FIG. 21



RO/

40/40

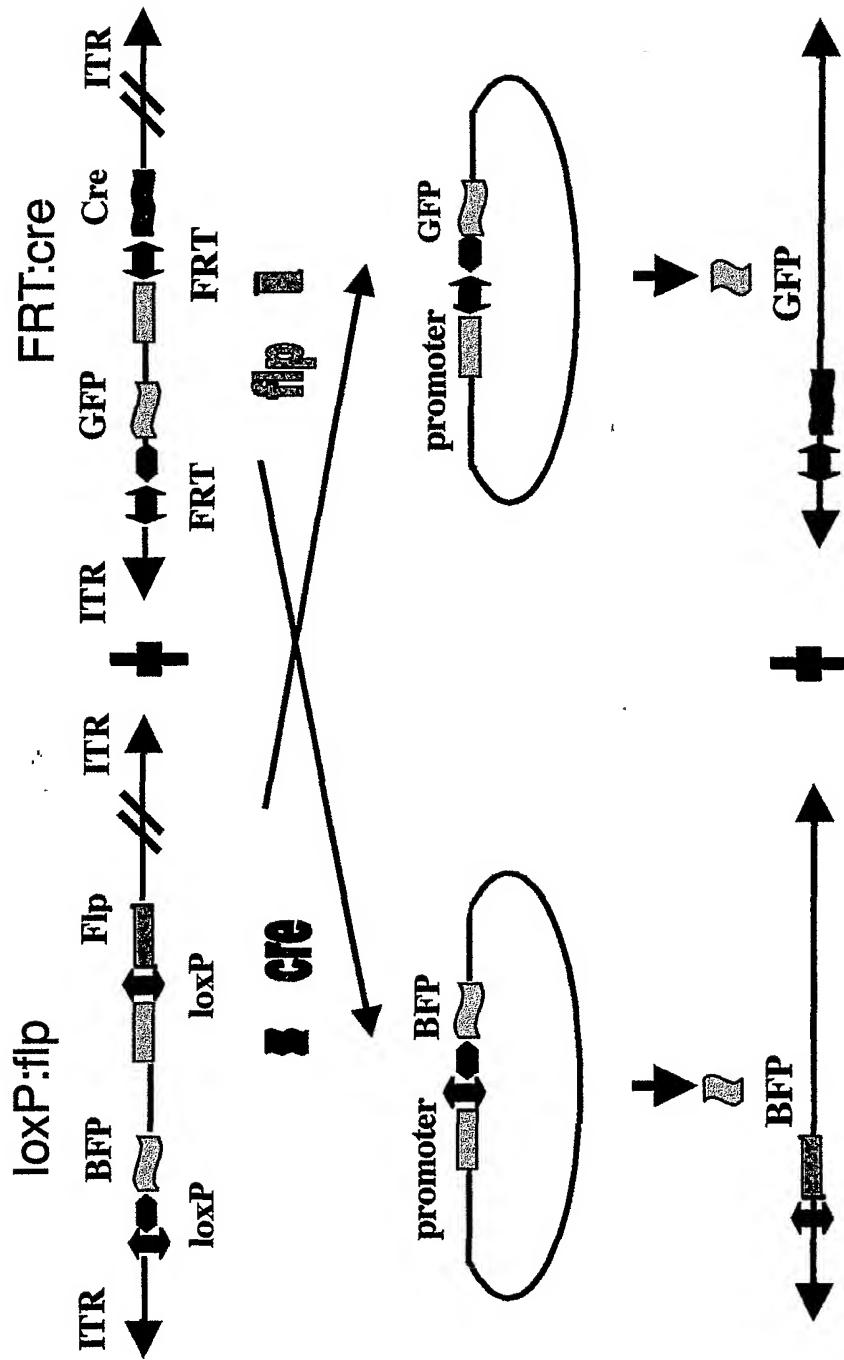


FIG. 22

SEQUENCE LISTING

<110> The General Hospital Corporation

<120> Self-rearranging DNA vectors

<130> 00786/352WO3

<150> US 60/231,053
<151> 2000-09-08

<150> US 60/246,904
<151> 2000-11-08

<160> 10

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 2341
<212> DNA
<213> Artificial Sequence

<220>

<223> derived from Adenovirus

<400> 1

gaatcggcca ggcgcaattc gattatcatc atcataaatat accttatttt ggattgaagc 60
caatatgata atgagggggt ggagtttg agctggcgac gggcgtggga acggggccgg 120
tgacgttaggt ttttagggcg agtaacttgc atgtattggg aattgttagtt tttttaaaat 180
ggaaagttac gtacgcggca tcgatgcgcg gatatcgcg gcggctagcg acatgagggt 240
gcccccgtatt cagtgtcgct gatttgtatt gtctgaagtt gtttttacgt taagttgatg 300
cagatcaatt aatacgtata ctgcgtcata attgatttatt tgacgtgggt tgatggccctc 360
cacgcacgtt gtgatatgtt gatgataatc attatcactt tacgggtcct ttccgggtat 420
ccgacaggtt acggggccgc gacctcgccg gtttcgcta ttatgaaaaa ttttccgggt 480
taaggcggtt ccgttcttct tcgtcataaac ttaatgtttt tattttaaaat accctctgaa 540
aagaaaggaa acgacaggtt ctgaaagcga ggcttttgg cctctgtcgt ttcccttc 600
tgttttggc cgtggaatga acaatggaa ttaacggatc caggccgcga gcaaaaggcc 660
agcaaaaggcga caggaaccgt aaaaaggccg cggttgcgtgc gttttccat aggctccgc 720
cccctgacga gcatcacaaa aatcaacgcg caagtcagag gtggcgaaac ccgacaggac 780
tataaagata ccaggcggtt cccctggaa gtcctctcg ggcgtctctt gttccgacc 840
tgccgcttac cggataacctt tccgccttc tccctcggg aagcgtggcg ctttctcata 900
gctcacgctg taggtatctc agttcggtt aggtcggtcg ctccaagctg ggctgtgtgc 960
acgaacccccc cggtcagccc gaccgctgcg ctttatccgg taactatcgt cttgagtcca 1020
accggtaag acacgactta tcgccactgg cagcagccac tggtaacagg attagcagag 1080
cgaggtatgt aggccgtgct acagagtct tgaagtgggt gcctaactac ggctacacta 1140
gaagaacagt atttggtata tgcgccttcg caaagccagt taccttcgga aaaagagttg 1200
gtagctctt atccggcaaa caaaccaccc ctggtagccg tggttttttt gtttgcagc 1260
agcagattac ggcgagaaaa aaaggatctc aagaagatcc ttgatcttt tctacgggg 1320
ctgacgctca gtggaaacgaa aactcacgtt aaggatttt ggtcatcaga ttatcaaaaa 1380
ggatcttcac cttagatcattt ttaaattaaa aatgaagttt taaatcaatc taaagtata 1440
attagtaaac ttggctgac agttaccaat gcttaatcag tgaggcacct atctcagcga 1500
tctgtctatt tcgttcatcc atagttgcct gactccccgt agttagata actacgata 1560
gggaggggctt accatccggc cccagtgctg caatgatacc gcgtgaccca cgctcaccgg 1620
ctcctgattt atcagcaata aaccagccag ccggaaagtgc cgagcgcaga agtggtcctg 1680
caactttatc cgcctccatc cagtcattt gttgttgcgg ggaagctaga gtaagtagtt 1740

<210> 2

<211> 34616

<212> DNA

<213> Artificial Sequence

<220>

<223> derived from Adenovirus

<400> 2

| | | | | | | |
|-------------|-------------|-------------|--------------|-------------|-------------|-----|
| gaatcggcca | gcgcgaaatta | actataacgg | tccttaaggta | gcgtcatcat | cataatatac | 60 |
| cttattttgg | attgaagcca | atatgataat | gaggggggtgg | agtttgtgac | gtggcgcggg | 120 |
| gcgtgggaac | ggggcggttg | acgttaggtt | tagggcggag | taacttgcatt | gtattggaa | 180 |
| ttttagttt | ttaaaaatgg | gaagttacgt | atcgtggaa | aacggaagtg | aagattttag | 240 |
| gaagttgtgg | gttttttggc | tttcgtttct | gggcgttaggt | tcgcgtgcgg | ttttctgggt | 300 |
| gttttttgtg | gacttaacc | gttacgtcat | tttttagtcc | tatataact | cgctctgtac | 360 |
| ttggcccttt | ttacactgtg | actgattgag | ctgggtccgt | gtcgagtggt | gttttttaat | 420 |
| aggtttttt | actggtaagg | ctgactgtt | tggctgcgc | tgttgaagcgc | ctgtatgtt | 480 |
| ttctggagcg | ggagggtgtgt | attttcctta | ggcaggagggg | tttttcaggt | gtttatgtt | 540 |
| ttttctctcc | tattaatttt | gttataccctc | ctatgggggc | tgtaatgtt | tctctacgcc | 600 |
| tgccggatgt | tatccccccg | ggcttattcg | gtcgctttt | agcaactgacc | gatgttaacc | 660 |
| aacctgtatgt | gttaccggag | tcttacattt | tgactccgga | catgaccggag | gaactgtcg | 720 |
| ttgtgttttt | taatcacggt | gaccagttt | tttacggtca | cgcggcgtat | gccgtatgtcc | 780 |
| gtcttatgtct | tataagggtt | gtttttcctg | ttgtaagaca | ggcttctaatt | gtttaaatgt | 840 |
| tttttttgtt | tattttattt | tgtgttttaat | gcaggaaccc | gcagacatgt | ttgagagaaaa | 900 |
| aatggtgtct | ttttctgtgg | tggttccgga | acttacctgc | ctttatctgc | atgagcatga | 960 |
| ctacgtatgt | cttgcttttt | tgcgcgaggc | tttgcctgtat | ttttttagca | gcacccgtca | 102 |
| ttttatatcg | ccgccccatgc | aacaagctt | cataggggt | acgctggta | gcatagctcc | 108 |
| gagttatgcgt | gtcataatca | gtgtgggttc | ttttgtcatg | gttccctggcg | gggaagtggc | 114 |
| cgcgcgggtc | cgtgcagacc | tgcacgatta | tgttcagctg | gccctgcgaa | gggacctacg | 120 |
| ggatcgggtt | atttttgtta | atgttccgct | tttgaatctt | atacagggtct | gtgaggaacc | 126 |
| tgaatttttg | caatcatgtat | tcgctgttt | aggctgaagg | tggagggcgc | tctggagcag | 132 |
| atttttacaa | tggccggact | taatattcgg | gatttgctt | gagacatatt | gataaggtgg | 138 |
| cgagataaaa | attattttggg | catggttgaa | ggtgctggaa | tgtttataga | ggagattcac | 144 |
| cctgaagggt | ttagccttta | cgtccacttg | gacgtggagg | cagtttgcct | tttggaaagcc | 150 |
| atttgtgcaac | atcttacaaa | tgccattatc | ttttcttcttgg | ctgttagagtt | tgaccacgccc | 156 |
| accggagggg | agcgcgttca | cttaatagat | cttcattttt | aggttttgg | taatcttttg | 162 |
| gaataaaaaaa | aaaaaaaaaca | tggttcttcc | agctctttcc | gtctccccc | tgtgtactc | 168 |
| gcagaacgaa | tgtgttaggt | ggctgggtgt | ggcttattct | gcgggtggtgg | atgttacatcg | 174 |
| ggcagcggcg | catgaaggag | tttacataga | acccgaagcc | agggggcgcc | tggatgttt | 180 |
| gagagatgtgg | atatactaca | actactacac | agagcgagct | aagcgacgag | accggagacg | 186 |
| cagatctgtt | tgtcacgccc | gcacccgtt | ttgcttcagg | aaatatgact | acgtccggcg | 192 |
| ttccattttgg | catgacacta | cgaccaacac | gatctcggtt | gtctcggcgc | actccgtaca | 198 |
| gtaggatcg | cctacacctt | tttgagacag | agacccgcgc | taccatactg | gaggatcatc | 204 |
| cgctctgtcc | cgaatgttaac | actttgacaa | tgccacaacgt | gagttacgtt | cgaggttttc | 210 |
| cctgcagtgt | gggatttacg | ctgattcagg | aatgggttgt | tccctggat | atggttctga | 216 |
| cgccggagga | gcttgtatc | ctgaggaagt | gtatgcacgt | gtgcctgtt | tgtgccaaca | 222 |

ttgatatatcat gacgagcatg atgatccatg gttacgagtc ctgggctctc cactgtcatt 2280
 gttccagtcc cggttccctg cagtgcata gccggcgcc ggttttggcc agctggtttta 2340
 ggatgggttgtt ggtatggcgcc atgtttaatc agagggttat atggtaccgg gaggtgggtga 2400
 attacaacat gccaaaagag gtaatgttta tgtccagcgt gtttatgagg ggtcgccact 2460
 taatctacct gcgcttggtt tatgtatggcc acgtgggtt tgggttcccc gccatgagct 2520
 ttgatatacag cgccttgac tttttttttt tgaacaatat tgggtgtctg tgctgcagg 2580
 actgtgctga tttaagttagt atcagggtgc gctgtgtgc ccggaggaca aggctctca 2640
 tgctgcgggc ggtgcataatc atcgctgagg agaccactgc catgttgat tcctgcagg 2700
 cggagcggcg gccgcagcag ttttttttgcg cgtctgtca gcaccaccgc cctatcctga 2760
 tgacgattatg tttttttttt cccatgttagt cgtggacttc cccctcgccg cccgttgagc 2820
 aaccgcaagt tggacacgcg cctgtggctc agcagctgga cagcgacatg aacttaagcg 2880
 agctgcccgg ggagtttattt aatatactg atgagcgttt ggtttttttt aaggccagcc 2940
 ggaatataac acctaagaat atgtctgtta cccatgatat gatgtttttt aaggccagcc 3000
 ggggagaaaag gactgtgtac tctgtgtt gggagggagg tggcagggtt aatactagg 3060
 ttctgtgagt ttgattaagg tacggtgatc aatataagct atgtgggtgtt ggggctatac 3120
 tactgaatga aaaatgactt gaaattttct gcaattgaaa aataaacacg ttgaaacata 3180
 acatgcaaca ggttcacgat tcttttatttcc tggccaatgt aggagaaggt gtaagagtt 3240
 gtagcaaaaag tttcagtgtt gtattttcca cttttccagg accatgtaaa agacatagag 3300
 taagtgccta cctcgctagt ttctgtggat tcactagtgc cattaagtgt aatggtaagt 3360
 atcataggtt tagttttatc accatgcaag taaacttgac tgacaatgtt attttagca 3420
 gtttgactttt gggtttttgg ataggctaga aggttaggca taaatccaaac tgcatttgt 3480
 tatggattttt cattagttga gttcccttattt ctaaagtttcc agtaatgtt ttaagtgaag 3540
 gagttctcca tttagaacacc gttttggctc aatctaagga atataactaac acttgcaacg 3600
 gtgcctgtca tggatgaaag atctccagat acagccaaag cagctacatg agcttagact 3660
 tgactcccac attttgcata aaccaaagat aatttgcagt cattatctga atgaattctg 3720
 cagtttaggag atgggtctgg gtttgcac agggtaagtt tgcattttttt tttttttcc 3780
 attgttaatgg cccctggatc gtcggaaatggt aaacccgctc caagttttttagt aatcatggca 3840
 ccgttttcat tggatcaat gccagagcca attttagttt tttttttttt gatatctgga 3900
 gactcagatg tggatgtatc aaactccaga ccctttccctg cattatagc tatggcagta 3960
 ttatcaaagt tttagtccact ggatttttt atgctaactt ccagttttt agtattgttt 4020
 gatgcattaa aaaggatata ggcctctgtt tagttatgt ccaagttatg agatgcattaa 4080
 atatacaggg gtccttgcctt cagtttaaga cgtatgtttt ttttggatc aaatgggtaa 4140
 tccacatcta gaatttaacaa gttgttattt atacgcattgc caccggccgt ttaatttcc 4200
 atgttgtttt atgaatcata accaatagct cctgcattttt tggttctaaag ggagttttgt 4260
 tcaacgggtga cacctggctc agtaactact gtttagtgtat cggagttttt tgctacttgc 4320
 aaaggaccgc ttatTTTAAAT tccttattttt ccattattta cattaaatagg atcttccatg 4380
 ttaatgccc Agctaccctgtt ggcagtagtt agcgggggttgc ttagcatttttcc atgcagttac agtaagggtt 4440
 tcgcgttgcac tgccagagag gggggctgtat tttttttttt ctagcttcc atctgacact 4500
 gtaatggggcc cttagttagc aatgtttttt ttggagtctt gcacgggtcag tggggcttgc 4560
 gactgtacgc taagagcgcc gcttagtaact atcagaggag cgggtttgc cactgttagg 4620
 ggccttgggg taatttgcata tgggtggggat gtttgcataac ttatgttttca ctttgggttgc 4680
 ttaatgttgtt ggtttttttt ggtttccggc ttgttcttagg 4740
 gtaagaccgc tgccctttt aagcgcatac atgcgttggg aggtgtccaa agttcggag 4800
 acggcttagag agagaactcc agggggactt tctttttttt cattgggtga aacaaatgg 4860
 ggggtttagaa agggccacatg tggaggcccg gtttctgtt cttatggata cacggggttt 4920
 aagggtgtttt cagacggctt ggcgcgtttc atctgcataa atatgaagat agtgggtgc 4980
 gagggacaag aacatgagga atttgcatttcc cttttttttt tttttttttt tttttttttt 5040
 aaaaggccgc tgagataccatg gttttttttt gttttttttt gttttttttt gttttttttt 5100
 gacaaagatt tgctgacttgc tttaatgttta gtaattttttt gttttttttt gttttttttt 5160
 tggatataaa tctcttataatc cacacatgtt tttaatataaa gttttttttt gttttttttt 5220
 ctgtatgggg aagtgcaggc agccctctgt ttctgcctgag tggcgggttgc cgtgtatagg 5280
 tttttctccc accataagca cttttttttt gttttttttt gttttttttt gttttttttt 5340
 gtttgcggta gttttttttt cttttttttt gttttttttt gttttttttt gttttttttt 5400
 tcttttacac tgggtgtttt gttttttttt gttttttttt gttttttttt gttttttttt 5460
 ggacgcccggcc ttgcgcctttt cttttttttt gttttttttt gttttttttt gttttttttt 5520
 atctagatgtt tttttttttt gttttttttt gttttttttt gttttttttt gttttttttt 5580
 gagaagggggg gcgaggctga ttgttgggg caataacccgtt ttgcagggtt atgacagcgg 5640
 gcactgggaa agtaggggtttt tttttttttt gttttttttt gttttttttt gttttttttt 5700

gatatatggc tagggcaaaa atggtactgc aaaaaaccat gacagagatg atggcgtata 5760
 accaggcttc tgacaaatcg ctctgtttgt ttagcagct gggaaatgttgc catatttgag 5820
 tgaatctgca gggaaatgtt cttttggag gcgcgtgaggt ttgggagcaa agcacaggta 5880
 gggcgcaaaa aatcagcaaa acaaaaatga cactccgtt cataattaaa gaattctgag 5940
 aagatcagct atagtcctgt ctctgttattt cgatgggtgc ctgaggtacg caatgcgcac 6000
 acaaaccaggc tcaatgaaact gatgaaaggc gatgactaca gtgacgaggc tgcatgag 6060
 gataagggtg acaaattccgt aaagcaggtt aactgtgaaa ggtggatgc aatctacttc 6120
 gatgtgagcg accgcggcca atgttagagca cgcacagaaa agcgaacaaa gggtaataa 6180
 tataagaact cgaggaatca tgtctcattt aatcatactg taaaagaaga gaacatgggt 6240
 tcagaccgtc caatctatga attttttcat ttagtgggtt gaggcacaatg ataggcctat 6300
 agatgggggg tctggcgcgt ctgcgttta ggcaacaaat aagccacata ataataaggc 6360
 aaacaaacat aagcgtatg gaaaaccacc acatgtccaa gctcgcccag tcattgacaa 6420
 aggcattgaac ttgggtaaa tttagggcag atgttagtcc ggttagcagt gtgtgcgt 6480
 agtccgttgtt gggcgcgtt gttgagccgg tcatactctgg agcaggcaag ctgaagctgg 6540
 gttttagtcaa atttgcgtt caggcgtt cagaaatcag gcgctaactg ccaggaaagt 6600
 ttgatttgaa ggttgggtt ataatttgc cccgcctggag catatcccac atagagtaaa 6660
 ttgtccaggg gaatacaagc aagcggaaaa tcaaggcatt ttcttttcat caataaaact 6720
 gcgtctgctt ttgtatttga gataaaagtaa ggtacatacc aaagcaagcg ctgtaataag 6780
 cagagcgggtg gaacaaaagg tgccagtgtt ctctaaacac ttttgtgggg gccacaactt 6840
 gtactgttttgc ttcattgtaca tggtaatatac gcacatttca taaaatggaa atttatacat 6900
 aaaagtttta cgattttcac ctggaaagac tgtgacatta tagtcgttag tgcacactgg 6960
 ctggccaaata gcatatacag cataacttgc aattttgtct ttgtggcga taataagctt 7020
 ttcatgttct gtgggtcatt ttataagaaat agtgcattca ttagcttctg atttaaatgt 7080
 aacattgcaat gctgggtcatt taaaactcaac ctttttggca gcgctgcaga ctgcccga 7140
 ggcgagcaag cttaaaatca tgtacctcat cttggatgtt gccccccagcg tttaaaaagc 7200
 tgacaataggc tacaaacgtg cgtgcagcag gcggcaaccc taaggcacag aagtgcgt 7260
 ataagaataa acagaattac aagagtaagg ataaccccgaa ccccaatttcc agaaaaattt 7320
 gacaagctt tagagttact tgaattgtc atataactttaa taaaaaaatc ccagcacccc 7380
 gaaaaatgct tttttgacct gagttccggg agttgagctc acctccctgtt ttggaaaaat 7440
 gggagtaatg tctggttacg ctcaggctgt aggtgtgggc gcagcaaccg gtgacgcact 7500
 cgtacgttcc cggcagggtga ggagggttgtt ggtgggttgtt ttttcttga cggtgttagtt 7560
 gaagccgaga aggttgtgtg gcaaaacttac ttcgtctcgc tggaaactgt tgtaaattac 7620
 aaatgaagag ccgttaaagt accaggttaag gtacttatttgc gcccgcgtt gcaaaccgga 7680
 ggtgagggtt gctttgggtc gctttgggtt ggtaaaaacg gtggcgttca caggatggcg 7740
 acaggagccc cagtagatttca taatttctgtt atttatttata ctcagcacag agatgacaac 7800
 aaagatctt atgtaatcca gggtaggac agttgcaaac cacggcgtaca acacaggac 7860
 cccgctcccg ctccacttagc agggggcgtc tggtaaactc ccgaatcagg ctacgtgtaa 7920
 gctctacctg ggtggtgagc cggacgcgtt ggcgcggggcc ctcgatatgc tttcggggca 7980
 attcaaagaatc acaaaaactca cccggagccgc gggcaaaagca ctttgtggcg cggcagtgg 8040
 cgagggtgtt caggcgcgtt cgcgtctgcgtt ctttgcgtt cattcgtcg tagccgtccc 8100
 ccgactctttt caccggcgtca aagttggggaa taaaactggtc cgggtgtt cggggaggc 8160
 cagaaaaaggg gttgaagtaa accgaaggca cgaactcctc aataaaattgt agagttccaa 8220
 tgcctccggc ggcggctcc gaggacgagg tctgcagatg taggatcgc tgacggggcg 8280
 taaatgaaga gcccgcagcg cccgcgtatc gaaatgtccc gtccggacgg agaccaagag 8340
 aggagctcac cgactcgtcg ttgagctgaa tacctcgccc tctgattttc aggtgagtt 8400
 taccctgccc gggcgaccgc accctgtgac gaaagccgcg cgcaagctgc gcccctgag 8460
 tagtcatctg aacttcggcc tggggcgttcc tgggaagatc cacagtgggtt ggagcgggac 8520
 tttcctggta caccaggggca gcccggccaa tacggggatt aaggttatta cgagggtgtgg 8580
 tggtaatagc cgcctgttgc aggagaattt ggttgcgtt ggcgcggatt ccgttgaccc 8640
 gggatatacat gtggggtccc ggcgtcatgt agtttatttgc ggtttagtag ttttggggcag 8700
 ctccagccgc aagtccatt tgggtgttgtt aactccacat gtagggcgtt ggaatttcct 8760
 tgctcataat ggcgtgtacg acaggtgtcg ggcgcgggtt tggccgttcc agatgacgta 8820
 gtttgcgcg ttaaatttga gaaaggggcgc gaaactagtc cttaaagatc agcgcgcagt 8880
 atttgctgaa gagagccctc gcgtcttcca gcgtgcgcgc aagctgatct tcgcttttgc 8940
 gatacaggca gctgcgggtt gggagcgcgca gagacctgtt ttttattttc agctcttgc 9000
 cttggccctt gctttgttgc aatatacgat acagagtggg aaaaatccat ttcttaagct 9060
 cgcgggtcga tacgggttgc ttggggcgc gacgcagcgc tcctcttcc gctgctggc 9120
 ccgctgttgcg tttcttgggc tttgtcagag tcttgcatac cggtcgccctt tgcttctgtg 9180

tgaccgctgc tggcgtgcc gctgccgtg ccgcgggtgc agtaggggtc gtagagatga 9240
cggttagtaat gcaggatgtt acggggaaag gccacgcccgt gatggtagag aagaaaagcgg 9300
cggcgaagg agatgttgc cccacagtct tgcaagcaag caactatggc gttcttgtc 9360
ccgcgccacg agcggtagcc ttggcgctgt tggtgtctt gggctaaccgg cggcggctgc 9420
ttagacttac cggccctggt tccagtgggt tccccatctac ggttgggtcg gccaacaggc 9480
agtgcggcgc ggcctgagg agcgagggtt gtagcgatgc tgggaacggc tgccaatttc 9540
tggggcgccc gcgaggggaa tgcgaccggag ggtgacggtg tttcgctgtga cacctcttc 9600
gcctcgaag ctgcgtctag gctgtcccgat tcttccatca tctcctctc ctcgtccaaa 9660
acccctctg cctgactgtc ccagtatttc tcctcgccg tgggtggcgg cggcggcagc 9720
tgcagcttct ttttgggtgc catcctggga agcaaggggcc cgccgctgtct gatagggtcg 9780
cggcggcggg gggattgggt tgagctcctc gccgactgg gggtccaggtaa accccccccg 9840
tcccttcgt agcagaaaact cttggcgggc tttgttcatg gcttgcattt ggccggcggg attggtcttc 9960
gtgcccctgg gtaatgacgc aggccgttaag ctccgcattt aatttgcgaa ggttaagccga 10020
gtagaaccta atctcggtgg cgtggtagtc ctcaaggtaa gacttttcgt caggcgaggg 10080
cgcccacagc cccggagtga gtttcaaccc cggagccgcg accctgcagc tcaaagggtac cgataatttgc actttcgcta 10140
caggagcgg tgccgggtgc ataggttgcgac ggcacagtga gctcacgtct tccatgtatgt cggagtggtt ggccgtcacc 10200
gacttcacagc caaagcggcg gagggttcattc acggtaacttgc gcaagggttag gtcagtttgcgaa ggttaagccga 10260
gcagtgaccg aaggcggcg gagggttcattc acggtaacttgc gcaagggttag gtcagtttgcgaa ggttaagccga 10320
aagcgcacag caggtggcgg gcagaatttgc tgaacgctctt gtcaggttttgcgaa ggttaagccga 10380
ttgcaacatg ctttgactgg tgaagtctgg cagaccctgtt gtcaggttttgcgaa ggttaagccga 10440
ttcggggaaag ataatgtccg ccaggtgcgc ggccacggag cgctcggttgcgaa ggttaagccga 10500
taggtcccttc aagtttgcg ttagcagctt ctgcagctcc ttttaggttgcg gctccctccag 10560
gcattgtcgc cacacgcccc tggccgttttgc caggtgttagt gtcaggttttgcgaa ggttaagccga 10620
gtcgccggacg tagtgcggc ggcgcctcgcctt cttgagcggtt gtcaggttttgcgaa ggttaagccga 10680
gaggcgggtt tcgtgcaaaaa ttccaaggtaa ggcagaccagg gtcaggttttgcgaa ggttaagccga 10740
aattttgcag gcttggcgca cgtagccctg gcaagggttgc gtcaggttttgcgaa ggttaagccga 10800
cttgcgttgc atctccgggtt cagcaaaaggaa cgcgtcgtatgc cactcaagcttgcgaa ggttaagccga 10860
aagcaactgcg gccatcatta gcttgcgtcg ctccctccaa gtcaggttttgcgaa ggttaagccga 10920
aagccagcgc gccagctgtc catcgccaaatgcggttgc gtcaggttttgcgaa ggttaagccga 10980
caagtttgcg tccctctccaa ggggtcggtc acggcgccacg atcagctcgc tcatgactgt 11040
gctcataacc ttgggggttgc ggttaagttgc cgggtaggca aagttttttgcgaa ggttaagccga 11100
gcgtttcagc acggcttaggc ggcgcgttgc accctcaagt tccaccagca ctccacagtgc 11160
actttcattt tcgctgtttt cttgttgcag accgttttgcg gtcaggttttgcgaa ggttaagccga 11220
aagaccctca aagatttttgc gcaacttcgtc gagcggaggcg atatcaggttgcgaa ggttaagccga 11280
ctgcccgaag gccagctgtc tgcgtcgatgc gtcaggttttgcgaa ggttaagccga 11340
cttgcagttt tggaaaaaaaatgttgcatggt ggcagacccatgcgaa ggttaagccga 11400
gtagaagtttggcgggttgggtcgca tgcgttttgcgaa ggttaagccga 11460
gcgcgggttag aacagggtggc gttcgtaggc aaggctgaca tccgctatgg cgagggggc 11520
atcgctgcgc tcttgcacatgc gtcgtcgatgc aatggcgac tggcgcttgcgaa gatgtttca 11580
cagcacgtcg tctccacatgc ttaggttagtc gccatgcctt tggccccccgcgacttgc 11640
ttcctcgttt gcttgcgtcg tgcgtcgatgc ttgcgttttgcgaa ggttaagccga 11700
atcctcgatgc ttttagcttgc aaaaacctgg gtcgtcgatgc ataatcaactt ctccttcctc 11760
aagcgggggtt gcttgcgtcg gcaagggttgc gtcgtcgatgc tggcgatcg 11820
ggttggcgaaac tcaaagggggg cggtaggttgc gtcgtcgatgc tgcgttttgcgaa ggttaagccga 11880
tttctgccta taggagaagg aaatggccag tcggaaagatgc gtcgtcgatgc tgggggggttgcgaa ggttaagccga 11940
cgagcgcggc gcggtgcggc cgcgtcgatgc accaaccatgcgaa ggttaagccga 12000
gcccgtcgccg ccgcctcccccc ggcgtcgcccg aaaaaaggcg aacccatgcgaa ggttaagccga 12060
cgaggacgaa gaagactcgatgc cacaagatgc gtcgtcgatgc tggcgccggg gatgtccgtac 12120
atcgacactcg acggcggatt tggccattgc gtcgtcgatgc aaaaagaagc gcccctctcc 12180
caagccccggc cggccgtccat ccccgagatgc gtcgtcgatgc agcgaggaaagc aagagaaga 12240
tgtggcgcta caaatgggttgc gtttgcgttgc gtcgtcgatgc ctaatcaactt acggcaaggc 12300
aggttaagcgc acggtgcggc ggctgaatgc gtcgtcgatgc gtcgtcgatgc tggcgccggg 12360
gcaagaggaa aaggaagatgc ccaatgcgatgc gtcgtcgatgc agcgttttgcgaa ggttaagccga 12420
gaggcgtccgc atcgtgtctg cgtggggagaa gggcatggatgc gtcgtcgatgc tgggggggttgcgaa ggttaagccga 12480
caagtaccac gtggataacgc atctaaaggc aacattcaactt gtcgtcgatgc tgggggggttgcgaa ggttaagccga 12540
agctctggcg gccgtatgcgaa agacctgggttgc aacaggaggatgc gtcgtcgatgc tgggggggttgcgaa ggttaagccga 12600
cttcaccacgc aacaagacacttgcgaa gtcgtcgatgc tgggggggttgcgaa ggttaagccga 12660

gtcgtttgc a gaggtaacct acaaggcacca cgagccccacg ggctgcgcgt tggcgtgc 12720
 ccgctgcgtc gagatcgaag gcgagctta gtgtctacac gggagcatta tgataaataa 12780
 ggagcacgtg attgaatgg atgtgacgag cggaaacccgg cagcgcgcgc tgaaggagca 12840
 gtcttagcaag gccaagatcg tgaagaacccg gtggggccga aatgtggtc agatctccaa 12900
 caccgacgc a aggtgcgtcg tgcacatgcgc ggcctgtccg gccaatcagt tttccggcaa 12960
 gtcttcgcgc atgttcttct ctgaaggcgc aaaggctcag gtggcttta agcagatcaa 13020
 ggcttcatg caggcgctgt atcctaacgc ccagaccggg cacggtcacc ttctgatgcc 13080
 actacggtgc gagtcaact caaaggctgg gcatgcaccc tttttggaa ggcagctacc 13140
 aaagttgact ccgttcgccc ttagcaacgc ggaggacact gacgcggatc tgatctccga 13200
 caagagcgtg ctggccagcg tcgacaccacc ggcgtgtata gtgttccagt gctgcaaccc 13260
 tgtgtatcg aactcgcgcg cgccaggccgg aggccccaaac tgctgactca agatatcgcc 13320
 gcccgcacccg ctaaacgcgt tggtgatggt ggcgcgcctg tggagtggaaa acttcaccga 13380
 gctgcccggg atggttgtgc ctgagttaa gtgggacact aaacaccagt atcgaacacgt 13440
 gtcctgcaca gtggcgcata ggcgtgcgc gcagaacccc tttgattttt aaacggcgca 13500
 gacggcaagg gtgggggta aataatcacc cgagagtgt aaaaaaaaaa catttgcctt 13560
 tattgaaagt gtctcctagt acattatttt tacatgtttt tcaagtgaca aaaagaagtg 13620
 gcgctcctaa tctgcgcact gtggctgcgg aagttagggcg agtggcgcctc cagaagctg 13680
 tagagctgtt cctgggtgcg acgcagggtg ggctgtaccc ggggactgtt aagcatggag 13740
 ttgggtaccc cggtaataag gtcatgggt ggggtgtat ccatgggagt ttggggccag 13800
 ttggcaaaagg cgtggagaaa catgcagcag aatagtccac aggccgcga gttgggcccc 13860
 tgcacgctt gggtgactt ttcagcgtt atacagcggg cgggggaaga agcaatggcg 13920
 ctacggcgca ggagtgactc gtactcaaacc tggtaaacct gtttgagtcg ttggtcagaa 13980
 aaggccaaagg gctcaaagag gtacgtgtt tttgagcgcg gtttccaggc aaaggccatc 14040
 cagtgtacgc ccccagtcgc ggcaccggcc gtattgacta tggcgcaggc gagtttgtt 14100
 ggagaaacaa agcctgaaaa ggccttgc ttaggtgccc aaaaatatgg cccacaacca 14160
 agatcttga caatggctt cagttcctgc tcactggagc ccatggccgc agctgtgtt 14220
 gatgtgtttt gcttctttt ttttgcggc ttgcccggc agaaggccgt ggcagggtac 14280
 acggtctcga tgacgcgcg gtgcggctgg tgcacacggc ccacgtcaaa gacttcaaac 14340
 aaaacataaa gaaggggtggg ctcgtccatg ggatccaccc caaaagtcat gtctagcgc 14400
 tggcgaggt tggcgtagag aagggtttgg cccaggtctg tgagtgcgc catggacata 14460
 aagttactgg agaatggat ggccaaagg gtgcgtacgc aaagaaacctt ttctgggta 14520
 atactgtcaa ccgcgtttt gcctttaggt gggtagggca ctttggccgg gtaaggctgt 14580
 ccctcgcgca tggtggggac gaggtgcct acgaatcctg agttgttatg ctggtaaga 14640
 attccaaacct gctgataactc ctgttattta gtatcgtca ccacttgcgc gctcatggc 14700
 tggaaagttt tgaagaacga gtacatgcgg tccttgcgc ttcttgcattt gtagaagccc 14760
 tggtagccaa tattgttagt ggcacaacatc tgcaccagga accagtctt ggtcatgtt 14820
 cactgagcta cgttgcgtcc ctccccgtca actgagcgtt taatctcaaa ctatggga 14880
 gtaagcaggc ggtcggtgcc cggccagcta acagaagagt caaaggtaat ggcacccccc 14940
 ttaaagggtgt gattaagata gaagggtccg tcaaggtatg gtatggagcc agatggatgt 15000
 tagtaagggt cgtagctga tcccgaggaa ggggtttctt ttgttccaa gcgtgtgaag 15060
 gcccaaccgc gaaatgtgc ccaaggcgc gatggatgg agatggccac gttggggcgc 15120
 ttggcggtt tggggatag catgttggc gcgaaaagggt agtcattaaa ggactggcgt 15180
 ttgggtgtcat ttctgagcat gggttccagc gtggaggccg ttttgcggc catggggaa 15240
 aagggtggcgt aaagacaaat gtcgtcaaac ttaatgttag ccccgtaac tctaagatcg 15300
 tttcccgagag agctctgcag aaccatgtt acatccttc tgaagtttca ttaatatgt 15360
 tatgagcctg gcaggaggag gagggttttta atggaaaaaa acttttgggg cactgtatg 15420
 tgaaggccaa cgtagccgc gtttcccaac aacatggagc gataacggag gcccgcat 15480
 cgggtgggtt taaaggatt aacgttgc tttttttttt tttttttttt tttttttttt 15540
 ttaatgttagc agtctacaag cccgggagcc accactcgct tttttttttt tttttttttt 15600
 ttgggggtgt cagatatttcc cacattgggt ggggttttta ttagcttgc tggcaggta 15660
 agcgcaatat tggagtaaaat gaaatttctc cataggttgg cttttttttt aatttccat 15720
 gcaaagggtgt tacccactcc tatttcattt cgtttgc aatgtttttttt tttttttttt 15780
 gtagtatctc cattatcgcc tgagccattt ccattttttt tttttttttt tttttttttt 15840
 gttaccccaa taccaccaag aggaaaacaa taatttggca attcatcctc agttccatgg 15900
 ttttcaatga ttcttaacatc tggatcatag ctgtctacag cctgatttca catagaaaaa 15960
 tatctggtcc tattttttttt tttttttttt tttttttttt tttttttttt 16020
 tcttgcaat ctaccacggc atttagctgc gatgcctgac cagcaagaac acccatgtt 16080
 ccagtgtgt tataatacat taggccaataa aaatttgc tttttttttt tttttttttt 16140

ctgtttggca tagattgttg acccaacata gcttagaat tttcatcacc tttccaggt 16200
 ttgttaagaca gatgtgtgc tggggtttcc atatttacat cttcaactgta caaaaccact 16260
 tttggtttag tagcatgcc ttgcgggtcg ttcaagagg tagtatttga gaagaattgc 16320
 aagtcaacat ttggaagagg cacccttt tcatccggaa ccagaacgga ttgaccacca 16380
 aaaggatttgc taggcctggc ataaagatcca tagcatggtt tcatgggagt tgggtttta 16440
 agcaacttc ctcctggcg attagcatca gtttcgttcc actgagatc gccaatttga 16500
 ggttctgggt gataggaagg atctgcgtat acaggtttag cttgtgttcc tgattgtct 16560
 gatcctattt gttagccgct tttttaatt gtttctccag acaaaggagc ctggcatag 16620
 acatgtgttt tcttagtagc ctgatctcg gctgtttgct cttcttcttc ctcttctca 16680
 tcttcatctt cctcttcttc atcctcggca actgcccggc cgctatcttc gtttgttcc 16740
 cactcacagg agttaggagc gcccgggaa gctagagcgt tggtaggcagt gccggagtag 16800
 ggcttaaaag taggccccct gtccagcacg ccggggatgt ctttgcacgt ggaagccata 16860
 tcaagcacac ggttgcacc cacagccagg gtgaaccgcg ctttgcacgt gtacgcggta 16920
 tcctcgcggc ccacagggtat gaaccgcagc gtcaaacgct gggaccggc tgggttacg 16980
 tcgtgcgtag gtgccaccgt ggggtttcta aacttggat tcaaggctgaa gtacgtctcg 17040
 gtggcgcggg caaaactgcac cagccggggg ctccaggtact ccatcttggaa aagcgggcgc 17160
 atgtgcatgt aagaccactg cggcatcatc gaaggggtag tacatgacac atacgacacg 17220
 acggcggctc agcagctcct ctggcggcga catggacgca cccccagacc cacgatgctg 17280
 tttagctattt agaagcatcg tcggcgcctc agggattgca tgcgtgcgg accgctggcg 17340
 ttcaactgtgc ttggccagggt gcaactggct acggggccca cggcggacgg ccctggcagc 17400
 gcacggcgcg gggacgcgcg gctagggggg gttacaacaa gcttacagg tgggttacgg 17460
 acagggttct gctgggtgtc agcggggggg ggcagggtca ccacggccgg tgggtgttgg 17520
 cccagactc cggtagccat gggcgcgatg ggacgggtgg tggcaggccc ttgttttagt 17580
 gcctcctcgt acgagggagg ctcatattt tgccatccaa gacactgtct ccacggccgg tggaggctcc 17640
 cgccgacgct tttcgcacg cccctcttggaa ttaatcttat ttgcactgc ctgggtggcc 17700
 tctacgggag ggcggggatc aagcttactg ttcatccaa ccaccttttgg tggaaaattt 17760
 aggtccacca ccccgtaat gccagaggcc aggccatcta tggtccaggc cttgctgcaca 17820
 tgctcttca acttgtccct cagcatcttg cctgtgtgc acagcgagcc ccagctgaag 17880
 tagttcttaa tgggtggaaacc gaaatttttta atgcgcctcc ttcccatgaa cgggcgcgag 17940
 gcgccaccgc tcatattgtc ggtggccata tcttgcctgtt ccacggccgg tgggttacgg 18000
 ccgtgtcgcg gggccagaga cgcaaaagtgg atgttttcca ttctacaaaa tagttacagg 18060
 accaagcgag cgtgagactc cagactttt attttggatt ttccacatgc aacttggttt 18120
 taatcagtgt ctctgcgcct gcaaggccac ggatgcaatt cggggcacgg cgccaatcgc 18180
 cgccgcgatc agtggaaataa ggagggggcag gataccgcgg cgcacatgcg ggtgcgacgc 18240
 gcgccgcgcg cgggtggcgc cacgacgcac gccggccgtc aggccgtggc cggccatgccc 18300
 cctcctacgg tgcattttc ctggaaatcc cggcacccggg aaacggaggc ggcaggtgag 18360
 ggcacatatct gcaagaacca caaagacccc cttttaaacgg atgctgggtt ggtagcgcgc 18420
 tggggcgcacc accagggtcc tgcctcttc gcgacccacc ctgcgcacgg aaatcggggc 18480
 cagcacgggc tggcgcacggc gacggggggcgc gcggttccca gtgggtggcc ggcgtcgggt 18540
 agtcgctcgt ttctggggc ggttaggtgtc gccacgatag ccgggggttag ggcgcgtgg 18600
 aggtatgttag gcatatttcgg cggtaggtgc gctggcggtg ccgtacttcc tggAACGGCG 18660
 cgggcgcggg gggctgaaa cgcgaaacat ccacgggtcc ttgtgcacct ccgttagaggt 18720
 ttggacgcgc gccgcagcgg cccctgcac cgcgcacatc gccaccggcc aggcaaccgg 18780
 ggacgtttgt gtctccatgc cctctgtggc agtggcaata ctatgtctac tgggtgtggg 18840
 tatctgaacg tccacggctc gcacgcccac tcccggtgcc acctgcttgc ttggccgcac 18900
 gcgacacctg ggctccagcc caggctccac ggtcattttt tccaaagacat ctccagtc 18960
 ctggcgctt ggtaccatca gtcgcacgtt ggggtccaa gtcaccagact cgcccttttag 19020
 gccgcgttt tcttggacgc gtcgcacgtt gggcagcacc tgctgcacgt tcacgggctt 19080
 taggttaggt gttgggttgc ctcgtccag cggcaacgcgc aacatgtctc tatgcccctt 19140
 tccgttaggca aactccccga ggcgcgtt ggcctgtca agcagggttcc cgtgcgcgt 19200
 cacctcatca tacacgcgc ttttaggtgc ggtggagcgc tcaccggcgc taaaaactac 19260
 ggtgggtccg ggtcgaaaa cacgtttac gcgtcgaccc ttccactgtc cccgcgcct 19320
 gggcgcggtt gcgtgcagca gttccaccc gtcgtcaagt tcatcatcat catcatctt 19380
 cttttctttt ttgacccgct ttagcttgc gggcttggaa tcctgctt cttcttcgg 19440
 gggccatag atctccggcg cgatgacccgt gagcatctct tctttgattt tgccgttgg 19500
 catagcttcg ttgcgcgcg cccgcgttgg atacatacaa cagtacgagt ctaagttagt 19560
 tttcttgca atcttagttgc gggggggggc ggtgcgcacg ggcacgcgc gggcgctaac 19620
 cgagtcgcgc acccagtaca cgttgcctt gcgaccctga gtcatagcactaatggccgc

| | | | | | | |
|-------------|-------------|-------------|-------------|-------------|-------------|-------|
| ggctgctgcg | gcggccgctc | gtcgccctgga | cctggggggc | acagtgacaa | tacccgcggc | 19680 |
| cagccattca | gcggcccgca | tggccgccc | tcggccggtg | cgacgtgcgc | gtttaagcag | 19740 |
| ggccggccgc | gcgcgttggg | cggcagtgcc | gggtcgccgg | cgggtggcgc | gtgtacgcgc | 19800 |
| cctccgcgt | cttttcat | tagcataacg | ccgggctccg | cgcaccacgg | tctgaatggc | 19860 |
| cgcgtccact | gtggacactg | gtggcggcgt | gggcgtgttag | ttgcgcgcct | cctccaccac | 19920 |
| cgcgtcaatg | gcgtcatcg | cggtggtgcg | cccagtgcgg | ccgcgtttgt | gcccgcggcc | 19980 |
| gggcgcgcgg | tagtgcgcgc | gcacgcgcac | tgggtgttgg | tcggagcgct | tcttgcccc | 20040 |
| gccaacatc | ttgcttggg | agcgcaggcc | ccagcctgtg | ttattgcgtt | gcgatataag | 20100 |
| gatggacatg | tttgctcaaa | aagtgcggct | cgataaggacg | cgcggcgaga | ctatgcggcag | 20160 |
| ggccttgtaa | acgtaggggc | aggtgcggcg | tctggcgtca | gtaatggtca | ctcgctggac | 20220 |
| tcctccgatg | ctgttgcgc | gcggtagcgt | cccgtgatct | gtgagagcag | gaacgttttc | 20280 |
| actgacggtg | gtgatggtgg | gggctggcgg | gcccgcggaaa | atctggttct | cgggaaagcgc | 20340 |
| attgaacacg | tgggtcagag | aggtaaactg | gcccgtggac | tgggagtaga | cggcctggtc | 20400 |
| gtttagaag | ctcttggagt | gcacgggcaa | cagctcggcg | cccaccacgg | gaaagttgct | 20460 |
| gatctggctc | gtggagcgg | aggtcacggg | gtctgcatac | atgtctggca | acgaccagta | 20520 |
| gacctgctcc | gagccgcagg | ttacgtcagg | agtgc当地 | agggtccatg | agcggatccc | 20580 |
| ggtctgaggg | tcgcccgt | tgtatgc当地 | gtaccagctg | cggtaactggg | tgaaggtgct | 20640 |
| gtcattgctt | attagggtgt | aactgcgtt | tttgcgtgtcc | tctgtcagg | gtttgatcac | 20700 |
| cgggttcttc | tgagggttct | cgacccctggg | ttgcgcagcg | ggggcggcag | cttctgcgc | 20760 |
| tgctcggcc | tcagcgcgt | tctccctccgc | ccgtgtggca | aagggtgtcgc | cgcgaatggc | 20820 |
| atatcggtt | atgtccatca | ccggctgc当地 | tgcgc当地 | ggccgcgtgg | agttcttctc | 20880 |
| cgcgcgcgt | ccactgttgc | tgcgc当地 | tgcgc当地 | ccgcctgtt | cgtgtcatac | 20940 |
| tttaagctt | gcctggtagg | cgtccacatc | caacagtgc | ggaatgttac | caccctccag | 21000 |
| gtcatcgtag | gtgatccaa | agccctcctg | gaagggttgc | cgcttgcgg | tgcccaacaa | 21060 |
| gttgcgtcagg | cggctgtggg | tgaagtccac | ccgc当地 | ggcagcaaaa | tgtatgtctg | 21120 |
| atgaaaggct | tcgtttgtat | ataccccaagg | catgacaaga | ccagtactg | ggtcaaacc | 21180 |
| cagtctgaag | ttgcgggtgt | caaactttac | ccgc当地 | tttccagaa | ccccgttctg | 21240 |
| cctgcccact | ttcaagtagt | gtccacatc | cgcttgc当地 | ataaggctca | tggtcatgg | 21300 |
| ctcggagtag | ttgcctcgg | gcagcgtgaa | ctccacccac | tcatattca | gtccacactg | 21360 |
| tttgcctta | gtaagcgagc | ggcacaccat | caccgc当地 | ttaaaacttat | tgttaaacat | 21420 |
| gaactcggtc | acatttggca | tgttggat | caggatgtt | ttcaggtcgc | cgc当地 | 21480 |
| cgaacgggtc | tcaagattga | tgttgc当地 | gttgc当地 | ccggggctgt | agtcattgtt | 21540 |
| ttgaatgacc | gtgggttagaa | agttgc当地 | gtcgttctgg | tagttcagg | atgc当地 | 21600 |
| cgttgacttgc | ttgtccacaa | ggtacacacg | ggtgtgtc | aatagggtg | ccaactcaga | 21660 |
| gtaacggatg | ctgttctcc | ccccggtagg | ccgc当地 | cgc当地 | caaacggcgg | 21720 |
| gtccagggg | gcatcgaa | gggaaacccag | cgccgc当地 | actggc当地 | cgctcaccac | 21780 |
| gctctcgtag | gagggaggag | gacccctcc | atacatcgcc | gcmc当地 | tactaagggg | 21840 |
| aatacaagaa | aaccaacgc | ccgtgc当地 | gccttgc当地 | gttttttatt | ttgcatcat | 21900 |
| ctttttttt | ttttttttaa | acatttctcc | ccagc当地 | gcaagggtc | gcaaacgggt | 21960 |
| tgcactccc | tcccaaattc | aggacgc当地 | tgttgc当地 | cgagtcatc | tctcccaca | 22020 |
| ccagaccccg | ctgacggc | tgccttgc | gacgggtgg | cgggc当地 | ccgggc当地 | 22080 |
| ccctgtgtc | ctgc当地 | gtcttccatc | tactcatctt | gtccactagg | ctctctatcc | 22140 |
| cgttgttggg | aatgcccgg | ggcagggtt | tttgc当地 | cggctgc当地 | agcgagttgt | 22200 |
| ttaggtactc | ctcctcgccc | agcaggc当地 | ggcgggtgtt | gcgagtg | gtaaaagacc | 22260 |
| ctatcaagct | tggaaatggg | ctactcgat | ctgaccgc当地 | ggccgc当地 | cctagatcg | 22320 |
| acaagctgct | tggcctgcgg | aagcttct | ttcgc当地 | cgccctcg | tgctcg | 22380 |
| gttgcactc | tagcagggtc | tgc当地 | gggaaaacac | gctgtcg | atgtcg | 22440 |
| agaggaatcc | atcgatcc | tgc当地 | caaattcccc | gtgtagaaa | ccaggggg | 22500 |
| gtagccagt | cgggttcaag | atggcatttgc | tgaatactc | gggttca | gcggccgc当地 | 22560 |
| gatgcagta | gtccattagg | cgattgataa | acggccgg | tgaggcatac | atgc当地 | 22620 |
| ccatgttgc | cgc当地 | tccagc当地 | cgctggc当地 | tacccgtc | cgcatcaggt | 22680 |
| taaggctcac | gctctgc当地 | acatagc当地 | agatgc当地 | ctcctcg | ttaaaactgt | 22740 |
| gcaacgaggg | gatctctgc | cgccgggtt | tcagc当地 | gttcagg | gc当地 | 22800 |
| tgcccggtc | ctcctgccc | agcgc当地 | tgacacttgc | aatctctt | aaagtatgt | 22860 |
| cgtccacatg | cgc当地 | atggcctc | ggtacagtg | cagcaagtg | cctaggtat | 22920 |
| tgtccgggg | cacgc当地 | ctgtcc | agggcg | tagcagc | aacaggc当地 | 22980 |
| agtgggg | cagcaagcta | gacacggc | cgccgtc | tgtggg | cgccacc | 23040 |
| acagccctg | caagttcttgc | aaagccttgc | tcagg | tttac | ggctctcg | 23100 |

tggctctggaa aaaatagtct ggcccggact ggtacacctc actttgcgggt gtctcagtc 23160
 ccattagccg cagtgcgctc acaaagtgg tgtagtcctc ctgtccccgc ggcacgttg 23220
 cgggctgtgt actcaggaag gcgtttagt caaccatggg gcccagggtt ccctgctgct 23280
 gcgcgcgctc acgctgcgcc acggcctcgcc gcacatcccc caccagccg tccaggttg 23340
 tctgcacgtt gccgtgttta taacgagcca cgcgtgaag cagcgcgtcg tagaccaggc 23400
 cggcctcatc gggccggatg gcccgtttt cggccagcgc gtttacgatc gccagcacct 23460
 tctcgctgt ggggtttgcg cgccggggca ccaccgctc cagaattgcg gagagccgg 23520
 tggctgctgg ctgctgccc aacgcgtcag gttacgcgc agtcagcgc atgatgcgg 23580
 ccatgacctg gcgcgcgtcg tccgtggagt taaggccgg a cggcgtggctc tgacgcggc 23640
 cccgcaccgc cgggtccgtt gcgtcttgcatacatgtc agaaaacatca ccgcttagta 23700
 ctgcgcgtcc tctggctgt actcatcgctc ctgcgtatata tcctccacgc cggcgcgtt 23760
 gccagcgcgc gccccgtggca cggccagccc aggtccggcc ccagctgcct ccagggcgc 23820
 tcggcttggg gcccagcgc ggtcagcgc cgcgtcaaaag taggactcgg cctctctatc 23880
 gcccgtggcc gtgccagcca gggcccttgc caggctgtgc atcagctcgc ggtcgctgag 23940
 ctgcgcgcgc cggctcacgc tcacggcctt gtggatgcgc tcgttgcgt aaacgcccag 24000
 gtcgtcgctc aaggtaagca cttcaacgc catgcgcgtg tagaaccctt cgtatcttac 24060
 ctccctgtct atgggaacgt aaggggatgt gtatatctt cgggcgtaaa acttgcggc 24120
 actgagcatg gaatagttaa tggccggccac ttgtcagcc aggctcaagc tgcgctctg 24180
 caccactatg ctctgcagaa tgtttatcaa atcgagcagc cagcggccct cgggctctac 24240
 tatgttttagc agcgcatccc tgaatgcctc gtttgcctc cttgtcgca ctataaggaa 24300
 cagctgcgc gatgagccggct tgctattttgg gtttgcctc cagcgcgttta caaagtccc 24360
 cagatgcattc agtccatag ccaccccttc ggcgcgcaca a cgcgcgtc cgtgggtgtt 24420
 aaagttttt tgaaagttaa tctctgggtt caccgtctgc tctgtacggg ttaccagggtc 24480
 gggccgcgc acgtgtgcgc gcgcgggact aatcccggtc aactcgcgc ctcgttgcgg 24540
 ctccctgcgc agcaaccgcg cgcgggttcag gccatgcgc aactcgcgc ctgcgtggaa 24600
 cttcgtatcc cgcacatcttc cgggcttcctc tccctcgcgg tcgcgaaaca gtttctgcgg 24660
 cggcacgtac gcctcgccgc tgcgtacgc cagctgcacc cttgggtgtc gctcaggaga 24720
 gggcgctcct agccgcgcgc ggcctcgcc ctccctccaa ggcgcgttcc gccaggtagt gccgggcccc 24780
 gcgcgcgcggg gtttcgtaat caccatctgc cgcgcgtca ggcgcgttcc ggcgcgttcc 24840
 tgacgcggta ggagaagggg aggggtccct gcatgtctgc cgcgttccctt gctcttgcgg 24900
 ctgctgagga gggggcgc tctgcgcag caccggatgc atctgggaaa agcaaaaaaaag 24960
 gggctgtcc ctgtttccgg aggaatttgc aagcgggtc ttgcgtacgc gggaggcaaa 25020
 ccccggttcg cgcagtcgg gccggcccgaa gactcgaacc gggggctctg cgactcaacc 25080
 cttggaaaat aaccctccgg ctacagggag cgagccactt aatgctttcg ctttccagcc 25140
 taaccgccta cgcgcgcgcgc ggccagtgcc caaaaaagct a cgcgcagcag cgcgcgcgc 25200
 tggaaaggaag cccaaaggag cgctccccccg ttgtctgcgc tcgcacaccc ggttgcaca 25260
 cgcggccgggt aaccgcatgg atcacggccg acggccggat cccgggttcg aaccccggtc 25320
 gtcgcgcatg atacccttgc gaatttattcc accagaccac ggaagagtgc cgcgttacag 25380
 gctctccctt tgcacgggtt cagacgtcaaa cgcgtcgc cgcgttcccg gccagagcgt 25440
 cccgaccatg gaggactttt tgcgcgtgc caacatctgg aaccgcgtcc ggcactttcc 25500
 gcgcgcctcc accaccgcg cggcgtac cttgtatgcg aggtacatct acggatatca 25560
 tcgcctttagt ttgaaagacc tcgcgcgcgcg agccccggcc accctacgc cgcgttccct 25620
 cccgcgcgcg cgcgcgcact ttttgggtggg atatcagttac ctgggtgcggc cttgcaacga 25680
 ctacgtctt gactcaagggtt cttactcgcc tctcaggatc accgagctct cgcagccggg 25740
 tcaccagacc gttaaactgtt cgcgttatggc caactgcact tacaccatca acacggcgc 25800
 ataccaccgc tttgtggaca tggatgactt ccagtcgtacc ctcacgcagg tgacgcaggc 25860
 catattagcc gagcgcgttgc tgcgcgcactt ggcctgtttt cagccgtatga ggggttcgg 25920
 ggtcacacgc atgggaggaa gagggcgc cctacggcca aactccgcgc cgcgcgttagc 25980
 gataagatgca agagatgcag gacaagagga aggagaagaa gaagtgcgg tagaaaggct 26040
 catgcaagac tactacaaag acctgcgcgc atgtcaaaac gaagcctggg gcatggccga 26100
 cccgcgcgcg attcagcagg cggcgcgcgc ggcacatggc cttctgtcga ccatccgcgc 26160
 tctcaagacc gcctacttta attacatcat cagcagcacc tccgcgcagaa acaaccccg 26220
 cccgcgcgcgc ctgcgcgcgc cgcgttccgc cagcgttccct tcgcactgtg actgggttaga 26280
 cgccttctc gagagggtttt cgcgttccgc cgcgttccct tcgcgttccgc cctctcggtt 26340
 cggagtaccc acacaacaat ttgttgcgtt cgcgttccgc ggcgtatccc tgccgcacgg 26400
 cagccccccgc ccaaccatca accgggacat gacggccgc gtcgttccac tgccgcgcgc 26460
 cgagaacggc cgcgcgcgtca cgcgcgcgtca gacggacat gacggccgc cgcggggaga tgatcgagc 26520
 ctttgcgtac cgcctcccg tgcgcgcgtc tgcgcgcgcgt gtcgttccct cccaccgc 26580

gccagaagaa gaagaagaag gggaggccct tatggaagag gagattgaag aagaagaggc 26640
 ccctgttagcc tttagcgcg aggtgcgcga cactgtcgcc gagctcatcc gtcttctgga 26700
 ggaggagta accgtgtcgg cgcgcaactc ccagtttgc aacttcgccc tggacttcta 26760
 cgaggccatg gagcgcttg aggcccttggg ggatataac gaatccacgt tgcgacgctg 26820
 gtttatgtac ttcttgcgtt cagaacacac cggcaccacc ctcaactacc tctttcagcg 26880
 cctcgaaac tacgcgttct tegccccgca cgtggagctc aatctcgccg aggtggatcat 26940
 gcgcgcccgc gatgccgaag ggggcgttgt ctacagccgc gtcttgcgaagc agggaggcct 27000
 caacgccttc tcgcagctca tggcccgcat ctccaacgcac ctcgcgcaca ccgtggagcg 27060
 agccggacgc ggagatctcc aggaggaaga gatcgagcag ttcatggccg aaatgccta 27120
 tcaagacaac tcaggagacg tgcaggagat ttgcgcctcg gccgcgtca acgacaccga 27180
 aattgattct gtcgaactct ctttcagggtt caagtcacc gggccgcgtc tttcacgca 27240
 gaggcgccag attcaggaga tcaaccgcgc cgtcgccgc ttgcgcgcga acctccgcgc 27300
 gcagcaccag ctccctggcc cgcgccgcgc cgacgtgccc ctgcgcgcgc tcccgccgg 27360
 tcccgagccc cccctaccc cggggggcccg cccgcgtcac cgcttttgc tgcatcatcc 27420
 aaggacaccc cgcggccca cgcgcgcgc cgcgttaccg tagtcgcgc gccccgatgc 27480
 ggcctttgc aagtcatcg tgcgcgcacc aaccagcccc tggaaatcag gtatcacctg 27540
 gacctagccc ggcgcctgac cccgcgtatgc gagtaaacc tgcaggagct cccgcctgac 27600
 ctgtcgccgc gggagctcca gaccatggac agctcccatc tgcgcatgt tgcatacaag 27660
 ctccgaccgc cgcgcgcgc catctggact ttggctcgcc gccgcgttgc ggtccgatcc 27720
 accataactc ccctcgagca gccagacggt caaggacaag cagccgaagt agaagaccac 27780
 cagccaaacc cgcgcgcgc ggggctcaaa ttccactct gcttccttgc ggcgcgtcg 27840
 caggtcaacc tgcgcgcgg tgcgcgcgc ggcgtactg cgcacgtttt 27900
 taaaaagcc agcgcgcgtt tgcgcgcgc cgcgcgcgc tctacttca ccacatcaac 27960
 agccactctt ccaacttgc gcccggagatc cgttctcc cgcgcgcgc ggcgcgcgc 28020
 accgagcgtc tctttgtcac ctacatgtt gacatgtt ggcgcgcgc ggcgcgcgc 28080
 aacgcgcgtc tgcccttcat gctggttatg aagttcgccg ggcgcgcgc ggcgcgcgc 28140
 gccgcgcgc gacgcgcgtt ggaccttggc tgggcgcgc ggcgcgcgc ggcgcgcgc 28200
 ttctactgca tcacccaga aaaaatggcc atagtcgcgc agttttaggac ctttcgcgc 28260
 cacctgcaaa tgctaatggc cctgtacccgc tggagctcat tcgtcgcttc caaccctcat 28320
 ctgcagact gggccctgtc agaacacggg ctcagctccc ctgaggagct cacctacgag 28380
 gaacttaaaa aattggccctc catcaagggc accccgcgc tcttggaaact ttacatcg 28440
 ggccacaaca tcaacggcctt cgacgcgcgc gtgcgcgcgc cccaggtat taacaaccgt 28500
 tccgagggtgc cgggaccctt cccgcgcgcgc cgcgcgcgc gggaaagata 28560
 ctttcaacg atgtcacccgc cggccctgcgc aacccgcgtt ccaaaaagcg cacggacttt 28620
 ttgtcttggg agcaggccgc atgcgcgcgc actgacttca aataccagta cctcaaagtc 28680
 atggttaggg acacccgtc gctcacccac acctcgctcc ggaaggccgc gcaggcatac 28740
 ggcgtacccgc tagaaaaaggg atgctgcgc taccaggccgc tcaaccaggat ctacatgcta 28800
 ggcttctacc gttcgaggc cgcgcgcgc cgcgcgcgc agtactggaa agaccgcgaa 28860
 gagtttgcgc tcaacccgcgc gctgtggaaa aaaaaggggc aggataagta tgacatcatc 28920
 aagaaaaccc tggactactg cggccctagac gtgcgcgcgc cgcgcgcgc ggcgcgcgc 28980
 ctgcgcgcact cctacgcctc cttcgccgt gacgcgcgcgc gtctcacaga cggccgcgc 29040
 aacgttccac agcgtccac catatcatcc aactcgcgc ccatcttcag gcagatagtc 29100
 ttccgaggcag agcaggccgc cccgcgcgcgc ctcgcgcgcgc acctcctcgc tccctcgac 29160
 gaactatacg attacgtgcgc cggccgcgcgc cgcgcgcgcgc gatgctaccc tacatatctt 29220
 ggaataactca gagagccctt ctacgtttac gacatttgcg gcatgtacgc cttcgccgc 29280
 acccaccctt cgcgcgcgcgc tcccccactc aaccatcatc agcgcgcgcgc tggccgcgc 29340
 gcatggcgcgc aggcgcgcgc cttgcgcgcgc actacttgcg cgcgcgcgcgc 29400
 ctggccgggg tctttaccgt ggacgcgcgcgc cccgcgcgcgc agacgcgcgcgc agacgcgcgc 29460
 cccgcattct gttcgccgc gggccgcgcgc ctctgcgcgc ccaacgcgcgc cctacgcgc 29520
 gaggttagcca ccagcggttgc cttgtcacc ctgcgcgcgcgc gcggttgcgcgc cgtgcacctg 29580
 gtggccgcgc agcgcgcgcgc cgttgcgcgcgc gatggccgcgc gcggttgcgcgc cgaataacgtg 29640
 cagctaaaca tcgcgcgcgc ggcgcgcgcgc gatgcgcgcgcgc gatgcgcgcgcgc gatgcgcgcgc 29700
 atcgccaaatg tgcgttccaa cggccctctac gggcgccgttgc ccaccaagct tgacaacaaa 29760
 aagattgtct tttctgcgc gatggacgcgc gccaccctca aaggcatcac cgcggccgc 29820
 gtgaatatac aatccctcctc gttttggaa actgacaatc ttagcgcaga agtcatgc 29880
 gctttgaga gggagtgactc acccccaacac gtcggccctcg cagacgcgc gtcggaaagag 29940
 agtgaggacg aacgcgcgcgc cacccttc tatagcgcgc cttcaggaaac acccggtcac 30000
 gtggccataca cctataaaacc aatcaccttc cttgtatgcgc gagaggccgcg catgtgtctt 30060

cacaccctgg agcgagtgga ccccctagtg gacaacgacc gctaccctc ccacttagcc 30120
 tccttcgtgc tggcctggac gggagccttc gtctcagagt ggtccgagtt tctatacgag 30180
 gaggaccgcg gaacaccgcg cgaggacagg cctctcaagt ctgtatacgg ggacacggac 30240
 agcctttcg tcaccgagcg tggacaccgg ctcatggaaa ccagaggtaa gaaacgcac 30300
 aaaaagcatg gggaaaacct gtttttgcac cccgaacggc cagagctcac ctggctcgtg 30360
 gaatgcgaga cctgctcgcc ggcctgcggc gggatgcct actccccgaa atcggattt 30420
 ctcgcgcaca agctctacgc ccttaaaagt ctgcactgcc cctcgtgcgg cgcctcctcc 30480
 aaggcaagc tgccgcacca gggccacgcc gcggaggggc tggactatga caccatggtc 30540
 aaatgctacc tggccgacgc gcagggcgaa gacccgcac gcttcagcac cagcaggacc 30600
 agcctaagc gcacccctggc cagcgcgcac cccggagcgc accccttcac cgtgaccagg 30660
 actacgctga cgaggaccct gcccggctgg aaagacatga ccctggcccg tctggacgag 30720
 caccgactac tgccgtacag cgaaagccgc cccaaccgc gaaacgagga gatatgctgg 30780
 atcgagatgc cgtagagcac gtgaccgagc tgtggaccg ccttgaactg ctttgtcaaa 30840
 cgctcaaaag catgcctacg gcccgcggcc tcaaaccgtt gaaaaactt gcttccttgc 30900
 aagaactgct atcgctggc ggcgagcggc ttctggcga tttggtcagg gaaaacatgc 30960
 aagtcaaggaa catgcttaac gaagtggccc ccctgctcag ggtatgcaccc agctgcagct 31020
 ctcttaacta ccagttgcag cccgtaatag gtgtgattt cgggcccacc ggctgcggta 31080
 agtcgcagct gctcaggaac ctgccttctt cccagctgtat cttcccttacc cccggaaacgg 31140
 ttttcttcat cggcccgccag gttagacatga tccccccatc tgaactcaaa gcgtgggaaa 31200
 tgcaaattctg tgagggttaac tacgccccctg ggccggatgg aaccattata cgcagtcgt 31260
 gcaccctccg cccgcgttt gtaaaaatgg cctatgcacca tctcatctg gaacacaact 31320
 atgacgttag tgatcccaga aatatctcg cccaggccgc cggccgtgg cccatttgca 31380
 tcattatggc cgaatggat gaaaaatctcg gaggtcacaa gggcgttcc aagtcttcc 31440
 acgcatttcc ttctaaagcta catgacaaat ttcccaagtg caccggatc actgtgctgg 31500
 tggittctgca caacatgaat ccccgaggg atatggctgg gaacatagcc aacctaaaaa 31560
 tacagtccaa gatgcacatc atatccccac gtatgcaccc atcccagctt aaccgcattt 31620
 taaaacactta caccaggc ctgccccctgg caatcagctt gctactgaaa gacattttt 31680
 ggcaccacgc ccagcgttcc tgctacgact ggatcatcta caacaccacc cccgagcatg 31740
 aagctctgca gtgggtctac ctccacccca gagacgggct tatgccccatg tatctgaaca 31800
 tccagagtca cctttaccac gtcctggaaa aaatacacag gaccctcaac gaccgagacc 31860
 gctggtcccg ggcctaccgc ggcgcacaaa cccctaaata aagacagcaa gacacttgct 31920
 tgatcaaaat ccaaacagag tctggttttt atttatgttt taaaccgcatt tgggggggaa 31980
 ggaaggcttc agggcagaaa cctgctggcg cagatccaaat agctgctgag aaacgcattt 32040
 aagttcccggt gtcacaaat ccaattgtgc caaaagagcc gtcaacttgt catgcggggc 32100
 ggtatgaacgg gaagctgcac tgcttgcac ggggctcagg aaagcaaaat cagtacaaat 32160
 cccgcgggcg gtggctgcag cggctgaagc ggcggcggag gctgcagtc ccaacggcgt 32220
 tccagacacg gtctcgtagg tcaaggtgt agagtttgcg ggcaggacgg ggcgaccatc 32280
 aatgctggag cccatcacat tctgacgcac cccggcccat gggggcatgc ggttgcataa 32340
 atatgagctc acaatgcacatcataaacat gttggcgctc atggcggccg ctgtgcacaa 32400
 acagatacaa aactacatga gaccccccatttataattt tttccaccc ttaagcccc 32460
 cccatcgatg gcaaacagct attatgggtt ttatgggtgc tagcgacatg aggttgcacc 32520
 gtattcgtgt tcgctgtttt gtattgtctt aagtttttt tacgttaatg tgatgcagat 32580
 caattaatac gatacctgcg tcataattga ttatttgcac tgggttgcatttgc gctccacgc 32640
 acgttgcgtat atgtatgtatc taatcattat cacttacgg gtcctttccg gtgatccgac 32700
 aggttacggg gcccgcaccc cgggggtttt cgctattttt gaaaattttc cgggttaagg 32760
 cgtttccgtt cttcttcgtc ataacttaat gttttttttt aaaataccct ctgaaaagaa 32820
 agggaaacgac aggtgtgaa agcgaggctt tttggcctct gtcgtttcc ttctctgttt 32880
 ttgtccgtgg aatgaacaat ggaagttAAC ggtatccaggc cgcgcacaaa aggccagcaa 32940
 aaggccagga accgtaaaaa ggccgcgtt gtcggctttt tccataggct cccgcacccct 33000
 gacgagcatc acaaaaaatca acgctcaagt cagaggtggc gaaacccgac aggactataa 33060
 agataccagg cgtttccccc tggaaagctcc ctcgtgcgtt ctcctgttcc gaccctgcgg 33120
 cttaccggat acctgtccgc ctttctccct tcgggaagcg tggcgcttcc tcatagctca 33180
 cgctgttagt atctcagttc ggtgttaggtc gttcgcttca agctggctg tgcacgaa 33240
 ccccccgttc agcccgaccg ctgcgcctt tccggtaact atcgtcttgc gtcaccccg 33300
 gtaagacacg acttatcgcc actggcagca gccactggta acaggattag cagagcgagg 33360
 tatgtaggcg gtgtacaga gttcttgcac tgggtggctt actacggctt cactagaaga 33420
 acagtatttg gtatctgcgc tctgccaaag ccagttaccc tcggaaaaag agttggtagc 33480
 tcttgatccg gcaaaacaaac caccgcgtt agcgggtggtt tttttgtttt caagcagcag 33540

attacgcgca gaaaaaaagg atctcaagaa gatccttga tcttttctac ggggtctgac 33600
 gctcagtgga acgaaaactc acgttaaggg atttttgtca tcagattatc aaaaaggatc 33660
 ttacccataga tccttttaaa ttaaaaatga agttttaaat caatctaaag tatatatgag 33720
 taaacctggt ctgacagtta ccaatgctt atcagtgagg cacctatctc akgatctgt 33780
 ctatccgtt catccatagt tgccctgactc cccgtatgt agataactac gatacgggg 33840
 ggcttaccat ccggccccag tgctgcaatg ataccggcg acccacgctc accggctcct 33900
 gatttatcag caataaaacca gccagccgga agtggcgagc gcagaagtgg tcctgcaact 33960
 ttatccgcct ccatccagtc tattagttt tgccggaaag ctagagtaag tagttcgcca 34020
 gttaatagtt ttcgcaacgt tggtgccatt gctacaggca tcgtgggtgc acgctcgctc 34080
 ttttgtatgg cttcattcag ctccggttcc caacgatcaa ggcgagttac atgatcccc 34140
 atgttgtgca aaaaagcggt tagtccttc ggtcctccga tagttgtcag aagtaagttg 34200
 gcccagtggt tatactcat gtttatggca gcactgcata attctcttac tgtcatgcca 34260
 tccgttaagat gctttctgt gactggtgag tattcaacca agaatacggg ataataccgc 34320
 gccacatago agaactttaa aagtgcctat cattggggaa cgttctcgg ggcggaaact 34380
 ctcaaggatc ttaccgctgt tgagatccag ttcatgtta cccactcgcg caccgaatgt 34440
 atcttctgca tcttttactt tcaccagctg ttctgggtga gcaaaaacag gaaggcaaaa 34500
 tgccgcaaaa aagggaataa gggcgacacg gaaatgttga atactcatac ttttccttt 34560
 tcaatattat tgaagcattt atcagggtt ttgtctcatc agcggataca tatttg 34616

<210> 3
 <211> 31672
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> derived from Adenovirus

<400> 3

gaatcgccca ggcgcaatta actataacgg tcctaaggta gcgtcatcat cataatataac 60
 cttattttgg attgaagcca atatgataat gaggggggtgg agtttgcac gtggcgccgg 120
 gcgtgggaac ggggcgggtg acgttaggtt tagggcggag taacttgcac gtattggaa 180
 ttgttagttt tttaaaatgg gaagttacgt atcgtggaa aacggaagtg aagatttgag 240
 gaagttgtgg gtttttggc ttctgttct gggcgtaggt tcgcgtgcgg ttttctgggt 300
 gtttttggc gactttaacc gttacgtcat tttttagtcc tataatataact cgctctgtac 360
 ttggccctttt ttacactgtg actgattgag ctgggtccgt gtcgagttgt gttttttaat 420
 aggtttttt actggtaagg ctgactgtt tggctgcgc tggaaagcg ctgtatgtt 480
 ttctggagcg ggaggggtgc attttgccta ggcaggaggg ttttgcgtt gtttatgtgt 540
 ttttctcctt tattatttt ttatataccctt ctatgggggc ttaatgttgc tctctacgc 600
 tgcgggtatg tattttttttt ggcttatttcg gtcgttccgtt agcactgacc gatgttaacc 660
 aacctgtatgtt gtttaccggat ttttacattt tgactccggat catgaccggat gaactgtcg 720
 tggtgctttt taatcacggat gaccgtttt tttacgggtca cgccggcatg gccgtatcc 780
 gtctttagtgc tataagggtt gtttttccgtt ttgttaagaca ggcttctaat gtttaaatgt 840
 tttttttttt tattttttttt ttgttttaat gcaggaaaccc gcagacatgt ttgagagaaa 900
 aatgggtgtct ttttctgtgg tggttccggat acttacctgc ctttatctgc atgagcatga 960
 ctacgatgtg cttgtttttt tgccggaggc tttgcgtat ttttgcgtt gacccctgca 1020
 ttttatatcg ccggccatgc aacaagctt cataggggtt acgctggat gcatacgctc 1080
 gagttatgcgt gtcataatca gtgtgggttc ttttgcgtt gttcctggcg gggaaagtggc 1140
 cgccgtggc cgtgcagacc tgcacgat ttttgcgtt gcccgtggc gggacctacg 1200
 ggatgcgggtt atttttgtt atgttccgtt ttgttaatctt atacaggatgt gtgaggaacc 1260
 tgaatttttgc ttttgcgtt gggatggcgc tctggagcag 1320
 attttacaa tggccggact taatattccg gatttgcgtt gagacatatt gataaggtgg 1380
 cgagatgaaa attatttggg catgggttga ggtgttggaa cagtttataga ggagattcac 1440
 cctgaagggtt ttagccctta cgtccactgt gacgtgagggtt ctttttttttgcgtt gtttgcct 1500
 atttgtgcac attttacaa tggccatttttgcgtt gtttgcgtt ctttttttttgcgtt gtttgcct 1560
 accggaggggg agcgcgttca cttaatagat ctttgcgtt gggatggcgc ttttttttttgcgtt gtttgcct 1620
 gaataaaaaaaa aaaaaaaaaaca tggcccttcc agcttccccc gtcctcccg ttttttttttgcgtt gtttgcct 1680

| | | | | | | |
|-------------|-------------|-------------|-------------|-------------|-------------|------|
| gtctttcacc | gcgtcaaagt | tgggaataaa | ctgggccggg | tagtggccgg | gagggtccaga | 5220 |
| aaaggggttg | aagtaaacgg | aaggcacgaa | ctccctaata | aattgttagag | ttccaatgac | 5280 |
| tccggagcgc | ggctcccgagg | acgaggctcg | cagagttagg | atcgccctgac | ggggcgtaaa | 5340 |
| tgaagagcgg | ccagccccgc | cgatctgaaa | tgtcccggtcc | ggacggagac | caagagagga | 5400 |
| gctcaccgac | tcgtcggtga | gctgaataacc | tcgcctctg | attttcaggat | gagttatacc | 5460 |
| ctggccgggc | gaccgcaccc | tgtgacgaaa | gcccccgcga | agctgcgccc | ctgagttagt | 5520 |
| catctgaact | tcggccctggg | cgtctctggg | aagtaccaca | gtggtgggag | cgggactttc | 5580 |
| ctgtacacc | agggcagcgg | gccaaactacg | gggatttaagg | ttattacgag | gtgtgggtgg | 5640 |
| aataggccgc | tgttcgagga | gaattcgggtt | tcgggtggcg | cggattccgt | tgaccgggaa | 5700 |
| tatcatgtgg | ggtccccgcgc | tcatgttagtt | tattcgggtt | gagtagtctt | ggcagctcc | 5760 |
| agccgcaagt | cccatttgtg | gctggtaact | ccacatgtag | ggcgtgggaa | tttccttgct | 5820 |
| cataatggcg | ctgacgacag | gtgctggcgc | cgggtgtggc | cgctggagat | gacgttagtt | 5880 |
| tcgcgcttaa | atttgagaaa | gggcgcgaaa | ctagtcctta | agagtcaagc | cgcagtattt | 5940 |
| gctgaagaga | gcctcccggt | cttccagcgt | gcccgaagc | tgatcttcgc | ttttgtgata | 6000 |
| caggcagctg | cgggtgaggg | agcgcagaga | cctgtttttt | attttcagct | cttgttcttgc | 6060 |
| gcccctgctt | tgttgaataa | tagcatacag | agtggaaaaa | atccctatttc | taagctcgcg | 6120 |
| ggtcgatacg | ggttcggtgg | gcccagacg | cagcgtcct | cctccctgctg | ctgcccggcgc | 6180 |
| tgtggattt | ttgggctttt | tcaagtcgtt | gctatccgg | cgcccttgct | tctgtgtgac | 6240 |
| cgctgctgtt | gctgccgctg | ccgctgcccgc | cgggtcagta | ggggctgttag | agatgacgg | 6300 |
| agtaatgcag | gatgttacgg | gggaaggcca | cggcgtgatg | gtagagaaga | aaggccgggg | 6360 |
| cgaaggagat | gttgcggcca | cagtcttgc | agcaagcaac | tatggcggtt | ttgtgcccgc | 6420 |
| gccacgagcg | gtagccttgg | cgtcttgg | gctttggc | taacggcggc | ggctgcttag | 6480 |
| acttacccgc | cctgggttcca | gtgggtgtcc | atctacgg | gggtcggcga | acaggcagt | 6540 |
| ccggcggcgc | ctgaggagcg | gagggtttag | cgatgtgg | aacggttgcc | aatttctggg | 6600 |
| gcggccggca | gggaaatgcg | accgagggtt | acgggtttt | gtctgacacc | tcttcggcct | 6660 |
| cggaagctt | gtctaggctg | tcccagtc | ccatcatctc | ctccctctcg | tccaaaacct | 6720 |
| cctctgcctg | actgtcccag | tattcctct | cgtcgtgg | ttggcggcggc | ggcagctgca | 6780 |
| gctctttttt | gggtgcccatt | ctgggaagca | aggggcccg | gctgtgtata | gggctgcggc | 6840 |
| ggcggggggg | ttgggtttag | ctcctcgcc | gactgggggt | ccaggtaaac | ccccctgtccc | 6900 |
| tttcgttagca | gaaactctt | gccccgtttt | ttgatggctt | gcaattggcc | aaggatgtgg | 6960 |
| ccctgggtaa | tgacgcaggc | gttaagctcc | gcatttggc | ggcgggattt | gtctcgtag | 7020 |
| aacctaattt | cgtggcggt | gtagtcctca | ggtacaaatt | tgcgaaggta | agccgacgtc | 7080 |
| cacagccccg | gagttagttt | caaccccg | gccgggact | tttcgttcagg | cgagggacc | 7140 |
| tgcagctcaa | aggtaccgt | aattttagtt | tcgctaagca | gttgcgaatt | gcagaccagg | 7200 |
| gagcgggtgcg | gggtgcata | gttgcagcga | cagtgcact | ccagtaggccc | gtcaccgc | 7260 |
| acgtcttcca | tgatgtcg | gtggtagg | aggtagttt | ctagctgcag | aaggtagcag | 7320 |
| tgaccccaaa | gcggcggagg | gcattcacgg | tactaatgg | gcacaaagtc | gctaggaagc | 7380 |
| gcacagcagg | tggcgggcag | atttcgtt | cgctctagga | taaagtttcct | aaagttttgc | 7440 |
| aacatgtttt | gactggtaa | gtctggcaga | ccctgttgc | gggttttaag | caggcggttc | 7500 |
| ggggaaatgg | tgtccggcc | gtgcggggcc | acggagcgt | cgttgaaggc | cgtccatagg | 7560 |
| tccttcagg | tttgcttttag | cagttctgc | agctcttta | gttgcgc | ctccaggcat | 7620 |
| tgctgccaca | cgcccatggc | cggttgc | gtgttagcaca | gaaataagta | aacgcagtc | 7680 |
| cggacgtgt | cgcggcgcgc | ctcgccctt | agcgttgg | gaagcacgtt | ttgcccggg | 7740 |
| cggtttcgt | gcaaaatttcc | aaggtagg | accagg | agagctccac | gttggaaatt | 7800 |
| ttgcaggcct | ggcgcacgt | gccctggc | aagggttagt | gcaacgtttc | ctctagctt | 7860 |
| cgctgcattt | ccgggtcagc | aaagaaccgc | tgcattgcact | caagctccac | ggtacaacagc | 7920 |
| actgcggcca | tcattagctt | gcgtcgctt | tccaagtcgg | caggctcg | cgtctcaagc | 7980 |
| cagcgcgc | gctgctcatc | gccaactgc | ggtaggcc | cctcggtt | ttcttgcaag | 8040 |
| tttgcatttt | tctccagggg | tcgtgcacgg | cgcacgatca | gctcgctcat | gactgtgtc | 8100 |
| ataaaccttgg | ggggtaggtt | aagtgcgg | taggcaaaat | gggtgaccc | gatgtgtgc | 8160 |
| ttcagcacgg | ctaggcgcgc | gttgcaccc | tcaagttcc | ccagcactcc | acagtactt | 8220 |
| tcattttcgc | tgtttctt | ttgcagagcg | tttgcgcgc | gtttctcg | gcgtccaaga | 8280 |
| ccctcaaaaga | tttttggc | tcgtcgagc | gaggcgat | caggatgac | agcgcctgc | 8340 |
| cgcacaggc | gctgcttgc | cgctcggt | cggttggc | ggcaggatag | gggtatctt | 8400 |
| cagttttgg | aaaagatgt | ataggtgg | agcacctt | gcacggcaaa | tacggggtag | 8460 |
| aagttgaggc | gcggggttgg | ctcgcatgt | ccgttttctt | ggcgttttgg | gggtacgcgc | 8520 |
| ggtgagaaca | gttggcg | gtaggcaagg | ctgacatcc | ctatggcg | ggcacatcg | 8580 |
| ctgcgtctt | gcaacgcgc | gcagataat | gcccacttgc | gctgcagat | cttcaacacag | 8640 |

acgtcgatcc ccacatctag gtagtcgcca tgcatttggc cccccggccc gacttgttcc 8700
 tcgtttgcct ctgcgtcgcc ctgggttgc tttttatccct ctgttggtac tgagcgatcc 8760
 tcgtcgatcc cgcttacaaa acctgggtcc tgctcgataa tcacttcctc ctccctcaaggc 8820
 ggggggtgcct cgacggggaa ggtggtaggc gcgttggcgg catcggtggaa ggcgggtgg 8880
 gcgactcaa agggggcggt taggctgtcc tccttcgtcg ctgactccat gatcttttc 8940
 tgcttatagg agaaggaaat ggcgcagtcc gaagaggagc agcgcgaaac caccggcgag 9000
 cgccggacgcg gtgcggcgcg acgtccacca accatggagg acgtgtcgcc cccgtcgccg 9060
 tcgcccgcgc ctccccgcgc gcccccaaaa aagcggctga ggcggcgctcg 9120
 gaccaagaag actcgacaca agatgcgcgtg gtgcgcgcga caccaggccc gcccgcac 9180
 acctcgacgg cgatttggc cattgcgtcc aaaaagaaaa agaagcgccc ctctcccaag 9240
 cccgagcgcc cgccatcccc agaggtgatc gtggacagcg aggaagaaag agaagatgtg 9300
 ggcgtacaaa tgggtgggtt cagcaaccca ccggtgctaa tcaagcacgg caagggaggt 9360
 aagcgacacgg tgccggcgct gaatgaagac gaccgagtgg cgccgggtat gcccgcac 9420
 gaggaaaaagg aagagtcacag tgaagcgaa agtgaagacg cggtgataaa cccgctgagc 9480
 ctgcccgtcg tgtctcggtg ggagaagggc atggaggctg cgcgcgcgtt gatggacaag 9540
 taccacgtgg ataacgatct aaaggcaaac ttcaagctac tgcctgacca atgtggaaagct 9600
 ctggcggcccg tatgcaagac ctggctaaac gaggaggcacc gcccgggtat gctgac 9660
 accagcaaca agaccttgt gacgatgtg gggcgattcc tgcaggcgta cctgcagtctg 9720
 tttgcagagg taacctacaa gcaccacgag cccacgggct ggcgttgc gctgcaccgc 9780
 tgcgtcgaga tcgaaggcga gcttaagtgt ctacacggga gcattatgtat aataaaggag 9840
 cacgtgattg aaatggatgt gacgagcga aacgggcgcg ggcgcgtgaa ggacgtct 9900
 agcaaggcga agatcgtaa gaaccgggtt ggccggaaatg tggtgacat ctccaaacacc 9960
 gacgcagggt gtcgcgtca tgacgcggcc tggcggccat atcagtttc cggcaagtct 10020
 tgccgcgtgt ctggcatgtt aaggcgcggaaatg gtcagggtgg cttttaagca gatcaaggct 10080
 ttcatgcagg cgctgtatcc taacgcggcc accgggcacg gtcaccccttct 10140
 cggtgcgagt gcaactcaaa gcctggcat gcacccttt tgggaaggca gctaccaaag 10200
 ttgactccgt tcgcccgtgaa acacgcggag gacctggacg cggatctgtat cttccgacaag 10260
 agcgtgctgg ccagcgtgca ccacccggcg ctgatagtgt tccagtgtg caaccctgtg 10320
 tattcgcaact cgcgcgcgca gggcgagggc cccaactcgcc acttcaagat atcggcgccc 10380
 gacctgctaa acgcgttggt gatggtgccg acgcgttggc gtgaaaactt caccgagctg 10440
 cccgcggatgg ttgtggctgaa gtttaagtgg agcactaaac accagtatcg caacgtgtcc 10500
 ctgcccgtgg cgcatagcga tgccgcggcag aacccttttgc atttttaaac ggcgcagacg 10560
 gcaagggtgg ggggtaaata atcaccggag agtgcataaa taaaacatt tgcctttatt 10620
 gaaagtgtct cctagatcat tatttttaca tggggatggcgtt gtcaccccttct 10680
 tcctaattctg cgcactgtgg ctgcggaaatg agggcgagtg ggcgcgtcagg aagctgtaga 10740
 gctgttcctg gttgcacgcg agggtgggtt gtacctgggg actgttaagc atggagttgg 10800
 gtaccccggt aataagggtt atgggtgggt tggatccat gggagttgg ggcgcagttgg 10860
 caaaggcgtg gagaacatcg cagcagaataa gtccacaggc ggcgcgttgc gggccctgca 10920
 cgctttgggt ggactttcc agcgttatac agcggcgggg ggaagaagca atggcgctac 10980
 ggcgcaggag tgactcgtac tcaaactgtt aaacctgtgc gactcggttgg ggcgcaggact 11040
 caaagggtctt aaagaggtagt catgttttgc agcgcgggtt ccaggcaag gccatccagt 11100
 gtacccccc agtctcgca cggccgtat gtcgtatggc gcaggcgagc ttgtgtggag 11160
 aaacaaagcc tggaaagcgc ttgtcatagg tgcccaaaaaa atatggccca caaccaagat 11220
 ctttgcataat ggcttcgtt ttcgtctcac tggagcccat ggcggcagct gttgttgatg 11280
 ttgttcgtt dttttatgtt gtggcggttc cggccggaaat gggcggtcgc agtacacgg 11340
 ttcgcgtatgc gccgcgggtc ggctgggtca cacggaccac gtcacccacttct 11400
 cataaagaag ggtggctcg tccatgggtt ccacccaaatg agtcatgtct agcgcgtgg 11460
 cggagttggc gtagagaagg ttttggccca ggtctgttag gtcgcctatg gacataaagt 11520
 tactggagaa tgggatgcgc caaagggtgc gatcgaaatc aaacttttc tgggtataac 11580
 tgcgtcgatcc ggtttggctt attagtgggtt agggcgttgc ggcgggtaa gcctgtccct 11640
 cgcgcgtt gggagcgagg tagcctacga atcctgttttgc tttatgttgg tgaagaattc 11700
 caacctgtcg atactcccttgc tattttgtat cgtcaaccac ttggccgctc atgggcttgg 11760
 agtttctgaa gaacgacttgc atgcggatcc ttttttttttgc tggaaatgtt ggtatggcc 11820
 agccaatattt gtagtggcc aacatctgcg ccaggaaatc gtccttggtc atgttgac 11880
 gagctacgtt gtagccctcc cccgtcaactg agcgtttaat ctcacactca ttgggagttaa 11940
 gcaggcggtc gttggccggc cagctaacatc aagagtcaaa ggtatggcc accttctttaa 12000
 aggtgttggatt aagatagaag gttccgtca ggtatggat ggcgcaggag taggtgttagt 12060
 aagggtcgta gcctgtatccc agggaaagggg tttccctttgtt cttcaagcgat gtcgaaaggccc 12120

aaccggaaaa tgctgcccag ttgcgcgatg gnatggagat gggcacgttg gtggcgttgg 12180
cggttatggg gtatagcatg ttggccgcgg aaaggtagtc attaaaggac tggtcgttgg 12240
tgtcatttct gacatggct tccagcgtgg aggccgtgtt gtggccatg gggagaagg 12300
tggcgtaaag acaaatgctg tcaaaactta tgctagcccc gtcaactcta agatcggttc 12360
ccagagagct ctgcagaacc atgttaacat ctttcctgaa gttccattca tatgtatatg 12420
agcctggcag gaggaggagg ttttaatgg caaaaaactt ttggggcacc tgaatgtgaa 12480
agggcacgta gcggccgtt cccaacaaca tggagcgata acggaggccc gcattgcgt 12540
ggtggtaaa ggattaaacg ttgtccatgt agtccagaga ccagcgcgccc ccaaggttaa 12600
tgtagcagtc tacaagcccg ggagccacca ctcgcttgcg catgtagtcg taggtgttgg 12660
ggttgtcaga tattccaca ttggggggt tgtattttag cttgtctggc aggtacagcg 12720
caatattgg aaaaaggaaa ttctccat gtttgcatt tagttaatt tccatgttgc 12780
agttgttacc cactcctatt tcattacgtg ttgaaaagt ttcatcttt gtcatgttag 12840
tatctcatt atgcctgaa ccattccat tagccatgt agttgtatag gtgtcaatgg 12900
ccccaaatacc cccaagagga aaacaataat ttggcaattc atccctgat ccatgggtt 12960
caatgattct aacatctgg aacatgtgt ttacagcctg attccacata gaaaaatatac 13020
tggttctatc acctatggaa tcaagcaaga ttgatagga cagctctgtg tttctgtctt 13080
gcaaatactac cacggcattt agtgcgtt agcaatgtaa atgttgcag 13140
tgctgttata atacattagg ccaataaaat tgcctctgaa agcaatgtaa ttggatgtt 13200
ttggcataga ttgttgacc aacatagctt tagaattttc atcacctttt ccaggtttgt 13260
aagacagatg tttgtctggg gtttccat ttacatcttc actgtacaaa accactttt 13320
gttttagtagc attgccttgc cggtgcgttca aagaggtatg atttgagaag aattgcaagt 13380
caaccttgg aagaggcacc ctttttcat ccggaaaccag aacggattga ccacaaaaag 13440
gattttagg cctggcataa gatccatagc atggttcat gggagttgtt ttttaagca 13500
ctctccctcc tggcgcatta gcatcagctt cggtccactg agattcgcca atttgggtt 13560
ctgggttata ggaaggatct gcgtatacag gtttagcttgc ttgtctgtc 13620
ctatTTTGTAG cccgcTTTTT gtaattgttt ctccagacaa aggagcctgg gcatagacat 13680
gtgttttctt agtagcctga tctcgagcgt tttgtcttc ttcttcctct tcttcatctt 13740
catctccctc ttcttcatcc tcggcaactg cccggccgct atttcgggtt tggtccact 13800
cacaggagt aggagcggcc ttggggacta gacgttgc ggcagtgcgg gagtagggct 13860
taaaagtagg cccctgtcc agcacccgc ggtatgtaaa gtacgtggaa gccatataca 13920
gcacacgggtt gtcacccaca gccagggtga accgcgtttt gtacgagttt gcggtatctt 13980
cgccgttccac agggatgaac cgcagcgtca acgcgtggaa cccgtctgtc gttacgtcgt 14040
gcgttagtgc caccgtggg tttcttaact tgttatttcg gctgaagtac gtcgttccat 14100
cgccggcaaa ctgcaccagg cggggctca ggtactccga gggttccctgg cccgagatgt 14160
gcatgtaaag ccactgcggc atcatcgaag gggtagccat cttggaaagc gggcgtacgg 14220
cggtctcagca gctcctctgg cggcgtacatg gacgcataca tgacacatac atgctgttca 14280
ctatTTAGAA gcatcgtcg gcttcaggg attgcacccc cagacccacg ctggcggcac 14340
gtgtgttctt ccagttgcca ctgggtacgg gccgtatcga tcgggacccg ggcagcacag 14400
ggcgcagggaa cgcgcggcta gggcggtta caacaacggc ggacggccct tgctggccca 14460
gtttctgtg ggtgtcagcg gggggaggca ggtccagcgt tacaggtgtt tttagtgcct 14520
gcactccggt agccatgggg gcgatggac ggttggtggg caggccttgc tcggcccg 14580
ctctcgtaa gggaggctca tctatttgcg tcaccagagt ttcttccctgt ggcgttccata 14640
gacgctttc gccacgcccc tctggagaca ctgtctccac gggcggtgg ggcgttccata 14700
cgggaggcg gggatcaagc ttactgttaa tcttattttg cactgcctgg ttggccagggt 14760
ccaccacccc gctaattgca gaggccaggc catctaccac cttttgttgg aaattttgtct 14820
ctttcaactt gtcctctcagc atctggcctg tgctgtgtt ccaggccttgc ctgcctatag 14880
tcttaatggt ggaacccgaa ttttaatgc cgccacacag cgagccccag ctgaaggcgc 14940
caccgctcat attgtctggt ccgatattctt gccgttccat gcataaacggg 15000
gtcgccgggc cagagacgca aagttgtatgt ctccatctt aaaaaatagt tacaggacca 15060
agcgagcgtg agactccaga ctttttattt tgattttcc acatgcataa tgggtttaat 15120
cagtgtctct ggcctgca ggcacggat gcaattccgg gcacggcc 15180
gcgatcagt gaaataaggag gggcaggata ccgcccgcga tgcgacgggt aatcgccgcg 15240
cgccgcgggtt ggtgcgcacg acgcattcccg cccgtcaggc cgtggccggc 15300
ctacggtgca ttcttcctcg gaatccggc accggaaac ggaggccggc ggtgaggggcc 15360
atatctgcaaa gaaccacaaa gaccggctt taaacgatgc tgggtggta ggcgcgttgc 15420
ggcagcacca gggtctgtcc tccttcgcga gcccatttcg gcacggaaat cggggccacg 15480
acgggctggc gacggcgaacg gcgccggccgg gttccagttt tgggtcgccg 15540
gctcgcttc tggggcggtt ggtgtagcca cgatagccgg gggtagggcgc gatggaaagga 15600

ttagggcat attcgggcag tagtgcgctg gcggtgccgt acttcctgga acggcgcccc 15660
 cgccgggggg ctgaaaacgcg aaacatccac gggccgttt gcacccctcg agagggtttt 15720
 gacgcggccg cagcggccgc ctgcaccgcg gcatctgcca ccggcgaggc aaccggggac 15780
 gtttgtct ccatgccctc tgtggcagtg gcaataactag tgctacttgtt ggtgggtatc 15840
 tgaacgtcca cggtctgcac gcccagtccc ggtgccacct gcttgattgg ccgcacgcg 15900
 acctcgggct ccagcccgagg ctccacgggt atttttcca agacatcttc cagtcgctgg 15960
 cgcttgggta ccatcagctc cacgggtgggt gccaagtcac cagactcg 16020
 cgctttctt cggacgggtc aagcgtggc agcacctgtc gcatgttcac ggcttttagg 16080
 ctaggtgtt ggttgcctc gtccagcggc aacccaaaca tgcctttagt ccgccttccg 16140
 taggcaaact ccccggaggc ctcgttggc tgctcaagca ggtcctcg 16200
 tcatacataca cgcgcttgc ggtgcgggtt gagcgttcac cggggctaaa aactacgggt 16260
 gtgccgggtc gcaaaacacg tcttacgcgt cgaccccttcc actgtacccg ccgcctgggc 16320
 gcggttgcgt gcagcgttc cacctcg 16380
 ttcttttga cccgctttag ctttcggggtt ttgtatcct gcttccctt cttcgggggg 16440
 ccatagatct ccggcgcgat gacccggagc atcttcttcc tgattttgcg ctggacata 16500
 gcttcgttgc ggcggccgcg cgctggatac atacaacagt acgagtctaa gtagttttt 16560
 ctgcataatct agttgcgcgg ggggggggtt cgacccggca cgccgaggcc gctaaccgag 16620
 tcgcgcaccc agtacacgtt gcccctgcga ccctgagtc tagactaat gggccggcgt 16680
 gctgcggcgg ccgctcg 16740
 cttcgagcgg cccgcattggc cgcccgctcg ccgggtgcac 16800
 gccggccgcg gttggggcgcg agtgcgggtt cggccggcggt ggcgacgtgc tacgcgcctc 16860
 cgcctctct tcattttagc ataacccccc gtcggcgca ccacggctcg aatggccgcg 16920
 tccactgtgg acactgtgg cggcgtggc gtgtatgtc 16980
 tcaatggcgt catcgacggt ggtgcggccca gtgcggccgc 17040
 gcgccgttagt gccgcgcac ggcactggg tggtggtcgg 17100
 aacatcttc ttggaaagcg caggccccag cctgtgttat tgctggcga tataaggatg 17160
 gacatgtttt ctcaaaaaagt gcgccgtcat aggacgcgcg 17220
 ttgttaacgt aggggcaggt gcggcgtctg gcgtcagtaa tggtcactcg 17280
 ccgatgtgt tgccgcagg tagcgtcccg tgatctgtga gaggcggaaac 17340
 acgggtgtga tgggggggc tggcggggcgc gccaaaatct 17400
 aacacgtggg tcagagaggt aaactggcgg atgagctggg agtagacggc 17460
 tagaaagctct tggagtgac gggcaacagc tcggggccca 17520
 tggctcg 17580
 tgcggtagt gacggaaatgt cacgggggtct tgcatcatgt 17640
 tgctccgagc cgcaggttac gtcaggagtg caaaggaggg 17700
 tgagggtcgc cgtagttgt tgcaaggta cagctcggt 17760
 ttgcttatta ggtgttaact gcgtttctg ctgtccctcg 17820
 ttcttcgtag gtttctcgac ctgggggtc gcaggggggg 17880
 tcggcctcag cgcgttctc ctccggccgt gtggcaaagg 17940
 tcgttcatgt cttccaccgg ctgcattgcc gcggctgccc 18000
 ccgctccac tgggttgcg gccgcctcg ccatccccgc 18060
 aagcttcgt ggtaggcg 18120
 tcgttagtga tcctaaagcc ctccctggaaag ggttgcgcgt 18180
 ctcaggcgcg tgggggtgaa gtccaccccg catcctggca 18240
 aagggttcgt ttgtatatac cccaggcatc acaagaccag tgactgggtc 18300
 ctgaagttgc ggggtcaaa ctttaccccg atgtcgctt 18360
 cccacttca agtagtgctc cacgatcg 18420
 gagtagttgc cctcgggcag cgtgaactcc acccactcat 18480
 tccttagtaa gcgagcgcga caccatcacc cgcgccttaa 18540
 tcgttcatat ttggcatgtt ggtatgcagg atggtttca 18600
 cggctcgaa gattgatgtt ctgtgtgc 18660
 atgaccgtgg ttagaaagtt gctgtgggtcg ttctgttagt 18720
 gacttgggtt ccacaaggta cacacgggtg gtgtcaata 18780
 cggatgtgtt ttctccccc ggtaggccgc aggtaccgcg 18840
 aggggagcat cgaaggggga acccagcgc 18900
 tcgttaggagg gaggaggacc ttccctcatac atcggccgcgc 18960
 caagaaaacc aacgctcggt gccatggcct tggtaggtt 19020
 tttttttttt ttttaaaaaca ttctcccccag cctggggcga 19080
 actccctccc aaatccagga cgctcg 19080

accccgctga cggtcgtgcc tttgacgacg ggtgggcggg cgccggccgg gcacatccct 19140
 gtgtcctgc gcatacgtct tccatctact catcttgc tccatctact catcttgc tccatctact 19200
 gttggaaat gccggaggca ggttctttc gcgctgcggc tgcagcagcg agttgttag 19260
 gtactcctcc tcgcccagca ggcgcggggc ggtgtgcga gtgctgtaa aagaccctat 19320
 caagcttggaa aatggctac tcgcatctga ccgcggggcc gcagcgccta gatcgacaa 19380
 gctgcttggc ctgcggaaagc ttcctttcg cagcgcggcc tctgcctgtc cgccgttgc 19440
 caactctagc agggtctgcg gttgcggggc aaacacgc tgctctatgt cgtcccaagag 19500
 gaatccatcg ttaccctcg gcacccatcg tcccccggtg tagaaaccag gggccggtag 19560
 ccagtgcggg ttcaagatgg cattgggtgaa atactcgggg ttccacggcgg ccgcgcgatg 19620
 caagtatgc attaggcgt tgataaaacgg ccgggtttag gtcatacatgc ccgggttccat 19680
 gttgcgcgcg gtcatgtcca ggcgcacgc gggcgttacc ccgtcgcgc gtcaggttaag 19740
 gctcacgctc tgctgcacat agcgcaagat ggcgcctcc tcgctgttta aactgtgcaa 19800
 cgaggggatc ttctggcgcc gggtggtcag caggtatgc aggttgcct ccaggctgccc 19860
 cgtgtcctcc tgccccagcg cgccgcgtgac acttgtaata tcctggaaaag tatgctcgtc 19920
 cacatgcgcc tgacctatgg ctcgcggta cagtgctcgc aagtgcacca ggtatgtgtc 19980
 ccgggacacg ctgcccactgt ccgtgaaggg cgctattagc agcagcaaca ggcgcgagtt 20040
 gggcgtcagc aagctagaca cggtcgcgcg gtcgcctgtg ggagccgc ccccccacag 20100
 cccctgcaag ttcttggaaag ctcggctcag gtttacggc tgccaggcctt gtctactgg 20160
 ctggaaaaaa tagtctggcc cggactggta cacccactt tgccgtgtc cagtcaccat 20220
 tagccgcagt ggcgcacaa agttgggtgta gtcctcctgt ccccgccgca cgttggcg 20280
 ctgtgtactc aggaaggcgt ttagtgcac acatggagccc aggttgcctt gctgctgcgc 20340
 ggcgcacgc tgccgcacgg ctcgcgcac atccccccacc agccggc tccagggtctg 20400
 cacgttgcgg ctgttgcac gggccacgcg ctgaagcgcg ggcgtgtaga ccaggccggc 20460
 ctcatcgccc cggatggccc tggtttcggc cagcgcgtt acgatgcgc gaccccttc 20520
 gtgcgtgggg ttgcgcgcg cgggaccac cgcttccaga attgcggaga gccgggtggc 20580
 ctggggctgc tgccggaaacg cgtcagggtt acgocgcgcg agcgcacatga tgccgttccat 20640
 gacttggcgc cagtcgtccg ttggatgaa gccggacggc tggctctgc ggcgcgcccc 20700
 cacccgcggg tccgttgcgt cttgcacat ctgatcagaa acatcaccgc ttgtactcg 20760
 ccgttccctg gtcgtactc atcgtcctcg tcatattcc tcacgcgc gacgttgc 20820
 ggcgcgcgg gtcgcacgc cagcccaagt ccggccccccag ctgcctccag ggcgcgtcgg 20880
 ctggggggcc accgcggc tccggccgcg tcaaaatggg actcggcctc tctatcgccc 20940
 ctggccgtgc cagccaggc ctttgcagg ctgtgcacat gtcgcgttgc gctcggttc 21000
 cggcccccgc tcacgcac cgccttgcg atgcgcgtc tgcgataaac gcccaggctc 21060
 tcgctcaagg taagcacctt caacgcacat cgcacatgcg acccctcgat cttacctcc 21120
 ttgtctatgg gaacgtaaagg ggtatggat atcttgcggg ctgttttttgc gtcggact 21180
 agcatggaaat agttaatggc ggcaccccttgc tcaagccaggc tcaagctgcg ctccctgcacc 21240
 actatgcctc gcaaatgtt tatcaaatacg agcagccaggc ggccttcggg ctctactatcg 21300
 ttttagcagcg catccctgaa tgcctcggtt tccctgtgt gtcgcactat aaggaacacgc 21360
 tgccgcacatgc gccggcttgcgat atttgggttt tgccgttgcg gcttacatgc gtcgttacaa gtcggccacaga 21420
 tgcatcagtc ctatagccac ctccctcgccg gccacaagcg tgcgttgcg gtcgcacatgc gtcgttacaa gtcggccacaga 21480
 cttttttgaa agttaatctc ctgggttccacc gtcgtgtcg gtcgcacatgc gtcgttacaa gtcggccacaga 21540
 gccgcacgt gtcgcgcgcg gggactaact cgggtccgcg cgtcgggcgc gtcgcacatgc gtcgttacaa gtcggccacaga 21600
 tcgcccacgc accgcgtcgcc gttcaggccca tgcgcacatgc gtcgcacatgc gtcgttacaa gtcggccacaga 21660
 cgatccccca ttcctcggg ctccctctcc tcgggttcgc acgtacgcct cgcgcgtgc acgttgcac gtcgcacatgc gtcgttacaa gtcggccacaga 21720
 ggcgcacatgc cgcgcgtgc acgttgcac gtcgcacatgc gtcgttacaa gtcggccacaga 21780
 gtccttagcc ggcgcaggcc ctgcgcctcc tccaaatgc gtcgcacatgc gtcgttacaa gtcggccacaga 21840
 cgggggggtt cgtaaatcacc atctggccgcg ggcgtcaggccg cggatgtgc ccctccgtac gtcgcacatgc gtcgttacaa gtcggccacaga 21900
 gcggtaggag aaggggaggg tgccctgcat gtcgttgc gtcgttgc gtcgcacatgc gtcgttacaa gtcggccacaga 21960
 tgaggagggg ggcgcacatgc cgcgcaggcc ggtatgcac gtcgcacatgc gtcgttacaa gtcggccacaga 22020
 tcgtccctgt ttccggagga atttgcac ggggttttc gtcgcacatgc gtcgttacaa gtcggccacaga 22080
 cgtcgcgcgc agtccggccg gcccggagact cgaaccgggg gtcctgcgac tcaacccttgc gtcgcacatgc gtcgttacaa gtcggccacaga 22140
 gaaaataacc ctccggctac aaggggaggc gcaacttaatgc gtcgcacatgc gtcgttacaa gtcggccacaga 22200
 cggttacgcgc ggcgcggcc agtggccaaa aaagctacgc gtcgcacatgc gtcgttacaa gtcggccacaga 22260
 aggaaggccaa aaggaggcgt ccccccgttgc gtcgcacatgc gtcgttacaa gtcggccacaga 22320
 ggcggtaacc gcatggatca cggcggacgg ccggatccgg gtcgcacatgc gtcgttacaa gtcggccacaga 22380
 gccatgatac ctttgcgaaat ttatccacca gaccacggaa gtcgcacatgc gtcgttacaa gtcggccacaga 22440
 tccttttgcg cggtctagag cgtcaacgcac tgcgcacgc tcaaccggcc gaggcgtcccg gtcgcacatgc gtcgttacaa gtcggccacaga 22500
 accatggagc acttttgcgc gtcgcacatgc gtcgttacaa gtcggccacaga 22560

gcctccacca ccgcccggg catcacctgg atgtccaggt acatctacgg atatcatcgc 22620
cttatgttgg aagacctcgcc cccggagcc cgggccaccc tacgctggcc cctctaccgc 22680
cagccggcgc cgcaactttt ggtggatat cagtacctgg tgccgacttg caacgactac 22740
gtctttgact caagggctta ctgcgtctc aggtacaccg agctctcgca gccgggtcac 22800
cagaccgtta actggtcgt tatggccaac tgcaattaca ccatcaacac gggcgcatac 22860
caccgcttg tgacatggg tgacttccag tctaccctca cgcaggtgca gcaggccata 22920
tttagccgagc gcgttgtcgc cgacctggcc ctgcttcagc cgatgagggg cttcgggtc 22980
acacgcatgg gaggaagagg ggcacccact cggccaaact cggccgccc cgtagcgata 23040
gatgcaagag atgcaggaca agaggaagga gaagaagaag tgccggtaga aaggctcatg 23100
caagactact acaaagaccc gcgcgcgtatgt caaaacgaag cctggggcat ggccgaccgc 23160
ctgcgcattc agcaggccgg acccaaggac atggtgctc tgcgaccat cccgcgtctc 23220
aagaccgcct actttaatta catcatcgac gcacacctcg ccagaaacaa cccgcaccgc 23280
caccgcgtc cggccggcc ggtgtcagc tacccctgcg actgtgactg gttagacg 23340
tttctcgaga gttttccga tccgtcgat gggactcgc tcaggtccct cgggtggga 23400
gtacctacac aacaatttgtt gagatgcatt gttagcggcc tattccctgcc gcacggcagc 23460
cccccgccaa cccataaccg ggacatgacg ggcggcgctc tccaaactgcg ccccgccgag 23520
aacggccgca cggtcaccga gaccatgcgc cgtcgccgcf gggagatgat cgagcgctt 23580
gtcgaccgca tcccggtcgcc cgcgtcgcc cgcgtgtcc cccctcccccc accgcccaca 23640
gaagaagaag aagaaggggg ggcctttagt gaagaggaga ttgaagaaga agaggcccct 23700
gtagcccttg agcgcgaggt gcgcgacact gtcggcgagc tcatccgtct tctggaggag 23760
gagttAACCG tgcggcgcc caactcccg ttttcaact tcggcggtgg cttctacgag 23820
gccatggagc gccttgaggc cttggggat atcaacgaat ccacgttgcg acgctgggt 23880
atgtacttct tcgtggcaga acacaccgc accaccctca actaccctt tcagcgctc 23940
cgaaaactacg cgcgttccgc cggcacgtg gagctcaatc tcgcgcagg 24000
gcccgcgtg cgaagggggg cgtggctac agccgcgtct ggaacgaggg aggcccta 24060
gccttcgtc agctcatggc cgcacatctcc aacgacctcg cgcgcaccgt ggagcgagcc 24120
ggacgcccgg atctccagga ggaagagatc gagcagttca tggccgaaat cgcctatcaa 24180
gacaactcag gagacgtgca ggagatttt cggcaggccg cgcgtcaacga caccgaaatt 24240
gattctgtcg aactctctt caggttcaag ctcacccggc cgcgtgtctt cgcgcagg 24300
cgccagattc agagatcaca cggccgcgtc gtgcgttgc ccacgttgcg cccgcgcag 24360
caccagctcc tggccgcgcg cggccgcgcac gtgcctctgc cccctctccc ggcgggtccc 24420
gagccccccc tacctccggg gggccgcgcg cgtcaccgt ttttagatgca tcatccaaagg 24480
acaccccccgc gggccaccgc cggccgcgcg gtaccgtat cgcgcgcggg ggtgcggcc 24540
tcttgcagaat catgcacgc gccaccaccc agccctggg aatcaggatg cactggacc 24600
tagcccgccg cctgaccgcg ctatgcgagg taaacctgca ggagctcccg cctgactgt 24660
cgccgcggg gctccagacc atggacagct cccatctgca cgatgttgc atcaagctcc 24720
gaccgcggcg cgcggacatc tggactttgg gctcgccg 24780
taactccct cgagcagcca gacgtcaag gacaaggcgc cgaagtagaa gaccaccagg 24840
caaaccggcc aggcggggg ctcaaattcc cactctgtt cctgtgcgc ggtcgtcagg 24900
tcaacctcgta caggatgta cagccgtgc accgctgca gtactgcgc cgtttttaca 24960
aaagccagca cgagtgttcg gcccgtgca gggacttta ctttaccac atcaacagcc 25020
actcctccaa ctggtggcg gagatccagt tttcccgat cggctcgcat cctcgaccc 25080
agcgtctctt tgtcacctac gatgtagaga cctataactt gatgggggccc tttggaaagc 25140
agctcggtcc ctcatgtct gttatgaagt tcggcggaga tgagcctctg gtcgaccacc 25200
cgcgagaccc acccggtggac ctggatggg accgctggg acaagaccgg cgcgaccacc 25260
actgcacatc cccagaaaaa atggccatag gtcgcacgtt taggacctt cctcatctt 25320
tgcaaatgct aatggcccg gacgtgtgg gtcattcgat cgcctccaaac cctcatctt 25380
cagactggc cctgtcgaaa cacgggtctc gtcctccgt ggcgtcacc tacgaggaac 25440
ttaaaaaaaaat gcccctccat aagggcaccc cgcgttctt ggaactttac atcggtggcc 25500
acaacatcaa cggcttcgac gagatgtgc tgcggccca ggttaattaaac aaccgttccg 25560
aggtgcggg acccttcggc atcacacgg actttatgc tcgcgcggg aagataactt 25620
tcaacatgtt caccatcgcc ctgcacccacc cgcgttccaa aaagcgcacg gacttttgc 25680
tctggagca gggcggtatgc gacgacactg attcaataa ccagtagc cctgtatgg 25740
tttagggacac ctttgcgtct acccacccat cgctccggaa ggcgcgcgcg gcatagcgc 25800
tacccgtaga aaagggtatgc tgcgcctacc agggcgtcaa ccagttctac atgctaggct 25860
cttaccgttc ggaggccgac gggttccga tccaagagta ctggaaagac cgcgaagagt 25920
ttgtccctaa cccgcgagctg tggaaaaaaa agggacagga taagtatgac atcatcaagg 25980
aaacccctgga ctactqcgcc ctagacgtgc aggtcaccgc cgagctggc aacaagctgc 26040

| | | | | | | |
|-------------|-------------|------------|-------------|-------------|-------------|-------|
| tgcgtggcaa | acagctatta | tgggttatt | gggtgcttagc | gacatgaggt | tgcggcgat | 29580 |
| tcagtgtcgc | tgattttgtat | tgtctgaagt | tgttttacg | ttaagttgat | gcagatcaat | 29640 |
| taatacgata | cctgcgtcat | aattgattat | ttgacgtgg | ttgatggcc | ccacgcacgt | 29700 |
| tgtgatatgt | agatgataat | cattatcact | ttacgggtcc | tttccgggt | tccgacagg | 29760 |
| tacggggcgg | cgaccctcg | ggttttcgct | atttatgaaa | atttccgg | ttaaggcg | 29820 |
| tccgttcttc | ttcgtcataa | cttaatgttt | ttattnaaaa | taccctctga | aaagaaaagga | 29880 |
| aacgacaggt | gctgaaagcg | aggcttttg | gcctctgtcg | tttccttct | ctgttttgt | 29940 |
| ccgtgaaatg | aacaatggaa | gttaacggat | ccaggcccg | agcaaaaaggc | cagcaaaagg | 30000 |
| ccaggaaccg | taaaaaggcc | gctgtgtgg | cgttttcca | taggctccgc | ccccctgacg | 30060 |
| agcatcacaa | aatcaacgc | tcaagtcaga | ggtggcgaaa | cccgcagagg | ctataaaagat | 30120 |
| accaggcggt | tcocccctgga | agctccctcg | tgcgtctcc | tgttccgacc | ctgcgcctt | 30180 |
| ccggatacc | gtccgcctt | ctcccttcgg | gaagcgtggc | gctttctcat | agctcacgc | 30240 |
| gttaggtatct | cagttcgggt | taggtcggt | gtcacaagct | gggctgtgt | cacgaacccc | 30300 |
| ccgttcagcc | cgaccgctgc | gccttatccg | gtaaactatcg | tctttagtcc | aacccggtaa | 30360 |
| gacacgactt | atcgccactg | gcagcagcc | ctggtaacag | gattagcaga | gcgaggat | 30420 |
| taggcgtgc | tacagagttc | ttgaagtgg | ggcctaacta | cggttacact | agaagaacag | 30480 |
| tatgtgtat | ctgcgtctg | ccaaagccag | ttaccttcgg | aaaaagagtt | ggtagctt | 30540 |
| gatccggcaa | acaaaaccacc | gctggtagcg | gtgggttttt | tgttgc | cagcagatta | 30600 |
| cgcgcagaaa | aaaaggatct | caagaagatc | ctttgatctt | ttctacgggg | tctgacgctc | 30660 |
| agtggAACGA | aaactcacgt | taagggattt | tggtcatcag | attatcaaaa | aggatcttca | 30720 |
| cctagatcct | tttaaattaa | aaatgaagtt | ttaaatcaat | ctaaagtata | tatgagtaaa | 30780 |
| cttggctcga | cagttaccaa | tgcttaatca | gtgaggcacc | tatctcagcg | atctgtctat | 30840 |
| ttcggttcatc | catagttgcc | tgactccccg | tagttagat | aactacgata | cgggagggct | 30900 |
| taccatccgg | ccccagtgct | gcaatgatac | cgcgtgaccc | acgctcaccg | gctcctgatt | 30960 |
| tatcagaat | aaaccagcca | gcccgaagt | ccgagcgcag | aagtggctct | gcaactttat | 31020 |
| ccgcctccat | ccagttatct | agtttgtgcc | gggaagctag | agtaagtat | tcgcgcgtt | 31080 |
| atagtttcg | caacgttgg | gccattgcta | caggcatcgt | gggtgcacgc | tcgtcg | 31140 |
| gtatggctc | attcagctcc | ggttcccaac | gatcaaggcg | agttacatga | tccccatgt | 31200 |
| tgtgcaaaaaa | agcgggttagc | tccttcggtc | ctccgatagt | tgtcagaagt | aagtggccg | 31260 |
| cagtgttatac | actcatgggt | atggcagcac | tgcatattc | tcttaactgtc | atgcacatccg | 31320 |
| taagatgctt | ttctgtgact | ggtgagtt | caaccaagaa | tacgggataa | taccgcgcca | 31380 |
| catagcagaa | ctttaaaagt | gctcatcatt | gggaaacgtt | cttcggggcg | aaaactctca | 31440 |
| aggatcttac | cgctgttgag | atccagttcg | atgtaaccca | ctcgccac | caagtgtatct | 31500 |
| tctgcatctt | ttactttcac | cagcgtttct | gggtgagcaa | aaacaggaag | gaaaaatgcc | 31560 |
| gcaaaaaagg | gaataagggc | gacacggaaa | tgttgaatac | tcataacttt | ccttttcaa | 31620 |
| tattattgaa | gcatttatca | gggttattgt | ctcatcagcg | gatacatatt | tg | 31672 |

```
<210> 4
<211> 30365
<212> DNA
<213> Artificial Sequence
```

<220>
<223> derived from Adenovirus

```

<400> 4
gaatcggcca ggcgcaatta actataacgg tccttaaggta gcgtcatcat cataaatatac 60
cttattttgg attgaagcca atatgataat gaggggggtgg agtttgtgac gtggcgccgg 120
gcgtgggaac ggggcgggtg acgttaggtt tagggcggag taacttgcat gtattggaa 180
ttgttagttt ttaaaaatgg gaagttacgt atcgtggaa aacggaagtg aagatttgag 240
gaagttgtgg gttttttggc ttctgttct gggcgttaggt tcgcgtgcgg ttttctgggt 300
gtttttgtg gacttttaacc gttacgtcat ttttttagtcc tatataact cgctctgtac 360
ttggccctt ttacactgtg actgattgag ctggtgccgt gtgcagtggt gtttttaat 420
aggttttttt actggtaagg ctgactgtta tggctgccgc tggaaagcgt ctgtatgtt 480
ttctggagcg ggaggggtgt atttgccta ggataacttc gtataatgtt tgctataacga 540
agttatggcg cgccagatct gtttgcacg cccgcacctg gttttgcctt aggaaatatg 600

```


gaccaagaga ggagctcacc gactcgctgt
ggtgaggattt accctgcggcc ggcgaccgca
cccctgagtt atgcatactga acttcggcct
gagcgggact ttccctggta accagggcag
gagggtgttgt gtaatagcc gcctgttcga
cggtgaccgg gatatcatg tggggcccg
cttggcagc tccagccgca agtcccattt
gaatttcctt gtcataatg gcgctgacga
gatgacgttag ttttcgcgt taaatttgag
gcccgcgacta ttgtcgaa agagcctccg
cgctttgtg atacaggcag ctgcgggtga
gctcttggc ttggccctg ctttggtaa
ttctaagctc gccccgtat acgggttcgt
ctgctgccgc cgctgtggat ttcttggct
gcttctgtgt gaccgctgt gtgtgtccg
tagagatgac gtagtaatg caggatgtta
agaaaaggc gggcaaggg gatgttggcc
ttcttgcgc cggccacga gcggtagcct
ggcggctgct tagacttacc gcccctgggt
cgaacaggca gtgcgggggg cgccctgagga
gccaatttct gggggccggg cgaggggaaat
acctcttcgg ctcggaaagc ttctgtctagg
tcgtccaaaa ctcctctgc ctgactgtcc
ggcggcagct gcaagtttctt ttgggtgcc
atagggtctc ggcgggggg ggattgggtt
aaccccccgt cccttcgtat gcagaaaactc
gccaaggatg tgccctggg taatgacgca
ttggtcttcg tagaacctaa tctctgggc
gtaagccgac gtcacagcc cggagttag
aggcgaggga cctcgagct caaagggtacc
attcgagacc agggagcggt gcgggggtca
gcccgtcaccg ctcacgttcc ccatgtatgc
cagaaggtag cagtgtaccc aaagccggg
gtcgctagga agcgacacgc aggtgggggg
cctaaagtt tcaacatgc ttgtactgg
aagcaggcgt tcggggaga taatgtccgc
ggccgtccat agtccctca agtttgcct
ctcctccagg cattgtctgcc acacggccat
gtaaacgcag tcgccccgt agtcggcg
gttttgcgg aggccgttt cgtcaaaat
cacgttgaa attttgcagg cctggcgcac
ttcctctagc ttgcgtctca tctccgggtc
cacggtaaca agcaactgcgg ccatcattag
gcccgtctca agccagcgcc ccagctgtc
ttgttcttcg aagtggcat ccctctccag
catgactgtg ctatataaccc tgggggttag
ctcgatgtg cgttcagca cggtaggc
tccacagtga ctttcattt cgcttgggt
gtcgcgtcca agaccctcaa agattttgg
gacagcggcc tggcgcaagg ccagctgtt
taggggtatc ttgcgtttt gaaaaaaagat
aaatacgggg tagaagttga ggcgcgggtt
ggggggtagc cgccgtgaga acaggtggcg
gaggggcaca tcgctgcgt ctgcacgc
atgcttcaac agcacgtcg ctccacatc
cccgacttgt tcctcggtt cctatgcgtc
tactgagcga tcctcgtagt ctgcgttac
ctcctccatca agcggggggtt ctcgacggg

ttagctgaaat acctcgccct ctgattttca 4140
ccctgtgacg aaagccgccc gcaagctcg 4200
gggcgtctct gggaaagtacc acagtgggtt 4260
cgggccaaact acggggatta agtttattac 4320
ggagaattcg gtttccgtgg ggcggattc 4380
cgctcatgtt gtttattcg gttgagtagt 4440
gtggctgtt actccacatg tagggcgtgg 4500
caggtgttgg cgccgggtgt ggccgttgg 4560
aaaggccgca aactactgtcc ttaagactca 4620
cgtcttcagg cggtgcggca agctgatctt 4680
gggagcgcag agacctgttt ttatatttca 4740
atatacgata cagagtggga aaaatcttat 4800
tggggccagg acgcagcgct ctcctctgt 4860
ttgtcagagt cttgtatcc ggtcgcctt 4920
ctggccgtgc cgccgggtca gtagggcgt 4980
cgggggagg ccacgcccgt atggtagaga 5040
ccacagtctt gcaagcaagc aactatggcg 5100
tggcgctgtt gttgtcttgc ggctaacggc 5160
ccagtggtgt cccatctacg gttgggtcg 5220
gcccgggtt tagcgatgt gggAACGGTT 5280
gcaaccgggg gtgacgggtgt ttcgtctgac 5340
ctgtcccagt cttccatcat ctcctctcc 5400
cagtatttct ctcgtccgt ggggtgcggc 5460
atccctggaa gcaagggccc ggggtgtctg 5520
gagctccctg ccggacttggg gttccaggta 5580
ttggcggtgt ttgttgcattgg gttcaatttg 5640
ggcggtaagc tccgcatttt gcccgggg 5700
gtgttagtcc tcaggtaaaa atttgcgaag 5760
tttcaacccc ggacccgggg acttttcgtc 5820
gataatttgc ttttcgttca gcaactgcga 5880
taggttgcag cgacagtgc actccagtag 5940
ggagtggttag gcaaggtagt tggctactgt 6000
agggcattca cggactttaa tggccaaaaa 6060
cagaatttctt gaaacgtctaa ggataaaatgt 6120
gaagtctggc agaccctgtt gcaagggtttt 6180
caggtgcgcg gccacggagc gctcgttcaa 6240
tagcagctt ctcagcttcc tttaggttgcg 6300
ggccgtttgc caggtgttagc acagaaataa 6360
cgccctcgccc ttgagctgtgg aatgaagcac 6420
tccaaggtag gagaccagg tgcagagctc 6480
gtagccctgg cgaaaagggtt agtgcacagt 6540
agcaaaagaa cgcgtcatgc actcaagctc 6600
cttgcgtcgc tcctccaatg cggcaggctc 6660
atcgccaact gcccgttaggc ctcctctcggt 6720
gggtcgttca cggccacgca tcagctcgct 6780
gttaagtgc ggttaggcaaa agtgggtgac 6840
cgcgttgcg ccctcaagt ccaccacgac 6900
ttgttgcaga ggcgttgcgg cgcgttctc 6960
cacttcgtcg agcgaggcga tatcaggat 7020
gtcccgctgg ctgcgggttgg cacggcaggag 7080
gtataggtt gcaagcaccctt cttggcactggc 7140
gggctcgcat gtgcctttt cttggcgtt 7200
ttcgttagca aggctgacat ccgtatggc 7260
gtcgcagata atggcgcact ggcgtcgcag 7320
taggttagtc ccatgcctt ggtcccccccg 7380
gtcctgggtt tgccttttat ctcctgttgg 7440
aaaacctggg tccgtctcgta taatcacttc 7500
gaaggtggta ggcgcgttgg cggcatcggt 7560

ggaggcggtg gtggcgaact caaagggggc ggttaggctg tcctcattct cgactgactc 7620
 catgatcttt ttctgcctat aggagaaggaa aatgccagt cgggaagagg agcagcgcga 7680
 aaccacccccc gagcgcggac gcgggtgcggc gcgcacgtcca ccaaccatgg aggacgtgtc 7740
 gtccccgtcg ccgtcgccgc cgcctcccg cgccgcggccca aaaaagcggc tgaggcggcg 7800
 tctcgagtcc gaggacgaag aagactcgta acaagatgcg ctggtgcgc gcacacccag 7860
 cccgcggcca tcgaccccgta cggcggattt ggcattgcg tccaaaaaga aaaagaaggc 7920
 cccctctccc aagcccgagc gcccggccatc cccagagggtg atcgtggaca gcgagggaga 7980
 aagagaagat gtggcgctac aaatgggtgg tttagcaac ccaccgggtgc taatcaagca 8040
 cggcaaggaa ggttaagcgca cgggtgcggcg gctgaatgaa gacgaccagg tggcgcgggg 8100
 tatgcgggacg caagaggaaa aggaaaggatc cagtgaaagcg gaaagtgaaa gcacgggtat 8160
 aaacccgcgtg agcctgcga tcgtgtctc gtgggagaag ggcattggagg ctgcgcgcgc 8220
 gtttatggac aagtaccacg tggataacga tctaaaggca aacttcaagc tactgcctga 8280
 ccaagtgaaa gctctggcgcc cctgtatgcgaa gacctggcta aacgaggagg accgcgggtt 8340
 gcagctgacc ttcaccaggca acaagacattt tgtgacgatg atggggcgat tcctgcaggc 8400
 gtacctgcag tcgtttgcag aggttaaccta caagcaccac gagccacacgg gctgcgcgtt 8460
 gtggctgcac cgctgcgctg agatcgaaagg cgagcttaag tgtctacacg ggagcattat 8520
 gataaataag gagcacgtga ttgaaatgga tgtgacgagc gaaaacgggc agcgcgcgc 8580
 gaaggaggcag tctagcaagg ccaagatcgta gaagaaccgg tggggccgaa atgtggtgca 8640
 gatctccaac accgacgcaa ggtgtcgct gcatgacgcg gcctgtccgg ccaatcagtt 8700
 ttccggcaag tcttgcggca tggatcttc tgaaggcgca aaggctcagg tggcttttaa 8760
 gcagatcaag gctttcatgc aggacgctgta tcctaaacgcg cagaccggc acggtcaccc 8820
 tctgatgcca ctacggtgcg agtgcacactc aaaggctggg catgcacccct ttttgggaaag 8880
 gcagactacca aagttgactc cgttcgccct gagcaacgcg gaggacctgg acgcggatct 8940
 gatctccgac aagagcgtgc tggccagcg gcacccaccc ggcgtatag tttccagtg 9000
 ctgcaacccct gtgtatcgca actcgcgcc gcagggcggc ggcggcaact gcaacttca 9060
 gatatcgccg cccgacacctgca aaacacgcgtt ggtatgggtc cgcagccgtt ggagtgaaa 9120
 cttcaccggc ctggccggaa tgggttggcc tgaggatggaa tgggactca aacaccagta 9180
 tcgcaacgtg tccctgcag tggcgcatag cgatgcgcgg cagaacccct ttgatttta 9240
 aacggcgcag acggcaaggaa tggggggtaa ataatcaccct gagagtgtac aaataaaaac 9300
 atttgccctt attgaaaggta tctcttagta cattattttt acatgtttt caagtgacaa 9360
 aaagaagtgg cgctcctaattt ctgcgcactg tggctgcgg gtagggcga gttgcgcgtt 9420
 aggaagctgt agagctttc ctgggtgcg cgcagggtgg gctgtacctg gggactgtta 9480
 agcatggagt tgggtacccc ggtataaagg ttcatgggtgg gttgtgatc catgggagtt 9540
 tggggccagt tggcaaaaggc gtggagaaac atgcagcaga atagtcacca ggcggccgag 9600
 ttggggccctt gcacgcctt ggtggacttt tccagcgtt tacagcggtc gggggaaagaa 9660
 gcaatggcgc tacggcgcag gagtactcg tactcaaact ggtaaacctg cttagtgcgt 9720
 tggtcagaaaa agccaaaggaa ctcaaaggagg tagcatgtttt tgagcgcgg gttccaggca 9780
 aaggccatcc agtgtacgc cccagctctcg cgacccggccg tattgactat ggcgcaggcg 9840
 agottgtgtg gagaacacaaa gcctggaaag cgcttgcgtcat aggtgccc aaatatatggc 9900
 ccacaaccaa gatcttgc aatggcttc agttctgtct cactggagcc catggcggca 9960
 gctgttgcgtt atgttgcctt cttctttat gttgtggcg tgcggccgaa gagggtgtt 10020
 cgcaggatcaca cggtctcgat gacgcgcggc tgccgtgtt gcatcaccc cgcacacgc 10080
 actctcaacaa aaacataaag aagggtgggc tcgtccatgg gatccaccc aaaagtcatg 10140
 tctagcgcgt gggcgagtt ggcgttagaga aggtttggc ccaggtgtt ggtgcgcgc 10200
 atggacataa agttactggaa gaatgggtatcg cgcacccagg tgcgatcgca aagaaacttt 10260
 ttotgggtaa tactgtcaac cgcgggtttt ggtaggcgc acgtcaaaag 10320
 taagcctgtc cctcgccat ggtgggagcg aggttagccta cgaatcctga gttgttatgc 10380
 tggtaagaa ttccaaacctg ctgatactcc ttgtatttt tagtgcgtt cacttgccgg 10440
 ctcatgggct ggaagtttctt gaaagacgag tacatgcgtt cttgttagct ttctggaaatg 10500
 tagaaaggccctt ggtagccat attgttagttt gccaacatct gcaccaggaa ccagtcctt 10560
 gtcatgttgc actgagctac gttgttagccc tccccgtcaa ctgagcgttt aatctcaaac 10620
 tcattgggag taagcaggcg tgcgttgcgg ggcacgtaa cagaagagtc aaaggtatg 10680
 gccacccctt taaagggtgt attaagatag aaggtccgtt caggtatgg tatggagcc 10740
 gaggtaggtgt agtaagggtc tgcgttgcgtt cccagggaa gggtttccctt tgcgttcaag 10800
 cgtgtgaagg cccaaaccgcg aaatgcgtcc cagttgcgcg atgggatgga gatgggcac 10860
 ttgggtggcggt tggcggttat ggggtatacg atgttggcg gggaaaggta gtcattaaag 10920
 gactggtcgt tgggtgtcatt tctgagcatg gcttccagcg tggaggccgt gttgtggcc 10980
 atggggaaaga aggtggcgta aagacaaatg ctgtcaaact taatgcttagc cccgtcaact 11040

ctaagatcg tttcccagaga gctctgcaga accatgttaa catccttcct gaagttccat 11100
 tcataatgtat atgagcctgg caggaggagg aggttttaa tggcaaaaaa cttttggggc 11160
 acctgaatgt gaaaggccac gtgcggccg tttcccaaca acatggagcg ataacggagg 11220
 cccgcattgc ggtgggtt aaagggattha acgttgtcca tgttagtccag agaccagcgc 11280
 gccccaaaggta taatgttagca gtctacaagc ccgggagcca ccactcgctt gtcatgttag 11340
 tcgttaggtgt tgggggtgtc agatattttcc acatgggtgg ggttgatatt tagttgtct 11400
 ggcaggta ggcgaatatt ggagtaaagg aaatttctcc ataggttggc atttaggtta 11460
 atttccatgg caaaggtagt acccactctt atttcattac gtgttgcaaa agtttcatct 11520
 tttgtccatgt tagtatctcc attatcgctt gagccattgc cattagcctt aatagcttga 11580
 taggtgtcag ttaccccaat acccccaaga ggaaaacaat aatttggcaa ttcattcctca 11640
 gttccatggt tttcaatgtat tctaacaatctt ggatcatagc tgcgtacagc ctgattccac 11700
 atagaaaaat atctgggttctt atcacctatg gaatcaagca agagttgata ggacagctct 11760
 gtgttctgt cttgcaaattc taccacggca tttagctgcg atgcctgacc agcaagaaca 11820
 cccatgttgc cagtgtgtt ataatacattt aggccaataa aattgtccctt gaaagcaatg 11880
 taattgggtc tggttggcat agattgttga cccaaacatag cttagaaattt ttcattcacct 11940
 tttccaggtt tctaagacag atgtgtgtctt ggggtttcca tatttacatc ttcaactgtac 12000
 aaaaccactt ttgggttagt agcattgcct tgccgtcgt tcaaagaggt agtatttgag 12060
 aagaattgca agtcaacattt tggaaagaggc acccctttt catccggAAC cagaacggat 12120
 tgaccaccaa aaggatttgtt aggccctggca taagatccat agcatggttt catggaggtt 12180
 gtttttttaa gcactctccc tcctgccgc ttagcatcag ctgcgttcca ctgagattcg 12240
 ccaattttagt gttctgggtt ataggaagga tctgcgtata caggtttagc ttgtgtttctt 12300
 gcattgtctg atccattttt tagcccccgtt tttgttaattt ttctccaga caaaggagcc 12360
 tggcataga catgtttttt cttagtagcc ttagtctcgat cgtttgtctt ttcttcttcc 12420
 tcttcattcat ctcttcatttca tcctccggaa ctgcgttcca gctatcttcg 12480
 gtttgcgttcc actcacagga gtttaggaggc cccttgggaa cttagacgtt gttaggcgtg 12540
 ccggagtagg gcttaaaaagt aggcccccttgc tccagcacgc cgccgtatgtc aaagtacgtg 12600
 gaagccatat caagcacacg gttgtcaccc acagccaggg tgaaccgcgc ttgtacgag 12660
 tacgcgttat cctcgcggc cacaggatg aaccgcagcg tcaaacgcgtt ggaccggct 12720
 gtgttacgt cgtgcgttagg tgccaccgtt gggtttctaa acttgcgttattt caggctgaag 12780
 tacgtctcggtt gggcgccggc aaactgcacc agcccggggc tcaggtactc cgaggcgtcc 12840
 tggcccgaga tggcatgttta agaccactgc ggcacatcatcg aagggttagc catcttgaa 12900
 agcgggcgcga cggcggctca gcaactcc tggggcgcac atggacgcac acatgacaca 12960
 tacgacacgt tagctatttta gaagcatcgt cggcgcttca gggattgcac ccccagaccc 13020
 acgatgcgtt tcagttgtgtt tgccagttt ccactggcta cggccgcacatcgac 13080
 ccgcgtggcgg cacggcgcag gacgcgcggg cttagggcggg ttacaacaac ggcggacggc 13140
 cctggcagca cagggttctg ctgggtgtca gccccgggag gcaagggtccag cgttacaggt 13200
 gtgtgctggc ccagcactcc ggttagccatg ggcgcgtatgg gacgggttggc gggcaggcc 13260
 tgcttagtgc cttccatcgta cgaggggaggc tcatcttattt ggcgtcaccag agtttcttc 13320
 ctgtcgggccc gggacgcctt ttcggccacgc ccctctggag acactgtctc cacggccgg 13380
 ggaggctctt ctacgggagg ggggggatca agctactgt taatcttattt ttgcactgc 13440
 tgggtggcca ggtccaccac cccgtaatgc ccagaggcca ggcacatctac caccttttgc 13500
 tggaaattttt gcttcttcaa ctgtgtccctt agcatctggc ctgtgctgtt gtccaggcc 13560
 ttgctgcccattt agttttaat ggtggaaaccg aaatttttaa tgccgccttca cagcgagcc 13620
 cagctgaagg cggccaccgtt catattgtgtt gtggcgatattt ctggccagttt tccatgaa 13680
 gggcgcgagc cgtgtcgccg gggccagagac gcaaaagttga tgcgttccat tctacaaaat 13740
 agttacagga ccaagcgagc gtgagactcc agactttta ttttgatattt tccacatgca 13800
 acttgcgtttt aatcagtgtc tctgcgcctt caaggccacg gatgcattt cgggcacggc 13860
 gccaatcgcc gggcgatca gtggaaataag gagggggcagg ataccggcgc gcatgcgacg 13920
 gtgcgacgcgc cggccggcc ggtgggtgc acgacgcattt cggccgtca gggccgtggcc 13980
 ggcacatggcc cttccatcggtt gcatcttcc tcggaaatccc ggcacccgggaa aacggaggcc 14040
 gcagggtgagg gccatatactg caagaaccac aaagaccggc ttttaaacga tgctgggttg 14100
 gtgcgcgtt gttggcagca ccagggtctt gcctccttcg cggccaccc tgcgcacgg 14160
 aatcggggccc agcacgggtt ggcgcacggc acggccggg cgggttccag tgggtggttcg 14220
 gcggtgggtt gtcgcgttcc ttctggggcc gtaggtgttag ccacgatagc cgggggtttagg 14280
 cgcgtatggaa ggtatgttaggg catattcggtt cagtagtgcgtt ctggcgggtcc cgtacttcct 14340
 ggaacggcgcg gggccggggg gggctgaaac gcaaaacatc cacgggtccg tttgcacccct 14400
 cgttagaggtt ttggacgcggc cgcgcacggc cgcctgcacc gcaaggcatctg ccaccggcc 14460
 ggcaaccggg gacgtttgtt tctccatgca ctctgtggca gtggcaatac tagtgctact 14520

ggtgggtgggt atctgaacgt ccacggctcg cacgcccagt cccggtgcca cctgcttgc 14580
 tggccgcacg cgAACCTCG gctccagccc aggcccacg gtcattttt ccaagacatc 14640
 ttccagtgcg tggcgcttgg gtaccatcg ctgcacgggt ggtgccaagt caccagactc 14700
 gcgctttagg ccgcgcctt cttcgacgg tgcaagcggt ggcagcacct gctgcagtgt 14760
 cacgggcctt aggctagggt ttgggttggc ctcgtccacg ggcaacgcca acatgtcc 14820
 atggcgctt ccgtaggcaa actccccgg ggcgtcggtt gcctgtcaa gcagggtcctc 14880
 gtgcggctac acctcatcat acacgcgcgt gtaggtgcgg gtggagcgct caccgggcgt 14940
 aaaaactacg gtggtgcgg gtcgcaaaac acgtcttacg cgtcgaccc tccactgtac 15000
 ccgcgcctg ggcgcgggtt cgtgcacggcgt tcgtcaagtt catcatcatc 15060
 atatatcttc tttttttt tgacccgtt tagtttgcg ggctttaat cctgtcttc 15120
 ctttttgcgg gggccataga tctccggcgc gatgacctgg tacataacaac agtacgagtc 15240
 gcgcttggac atagcttcgt tgcgcgcgcg cgcgcgttgg aaaaactacg gtggtgcgg 15300
 taagtagtt tttcttgcaa tctagttgcg cggggggcgg gcccctaacc gagtcgcgc 15360
 aatggccgcg gctgctgcgg cggccgcgt tcgcctggac cgcgcgttggc gacgcgc 15420
 acccgccgc accgttcgag cggccgcgttggc ggcgcgttggc gtcgcacgg gacgtgcgc 15480
 gttaagcagg gccgcgcgcg cgcgttggc ggcgcgttggc tgctacgcgc ctccgcgc 15540
 ctgaatggcc gcgccactg tggacactgg tggcgccgtt cggccgttggc gacataacgc 15600
 ctccaccacc gcgtaatgg cgtcatcgac ggtgggtgcgc cggccgttggc gacataacgc 15660
 cgcgcggcc ggcgcgcgttggc gacgcgc 15720
 ctttgcggcc ggcgcgcgttggc gacgcgc 15780
 ctttgcggcc ggcgcgcgttggc gacgcgc 15840
 cgatataagg atggacatgt tgctcaaaa agtgcggc 15900
 tatgcccagg gccttgcgttggc ggtgcggc 15960
 tcgctggact cctccatgc tggtgcgcag cggtagcgtc 16020
 aacgtttca ctgcacgggt tgatgggtgg ggcgcggcc 16080
 gggaaagcga ttgaacacgt gggtcagaga ggtaaactgg 16140
 ggcctggcgttggc ttgtagaagc tcttggaggtt gacgggcac 16200
 aaagttgtcg atctggctcg tggagcggaa ggtcacgggg 16260
 cgaccatgtt acctgcgttggc agccgcgttggc gacgcgc 16320
 gcgatcccg gtctgagggtt cggcgtagt gttatgc 16380
 gaagggtctgtc tcattgttta ttaggttgcgactgcgtt 16440
 tttgatcacc ggtttcttgc gaggcttcgac gacccgggtt 16500
 ttctgcgcgtt gcctcggcgtt cagcgcgtt ctcctccgc 16560
 gcaaatggca tgatgttca tgcctccac cggctgcatt 16620
 gttctcttcc ggcgcgttgc cactgttgcgactgcgtt 16680
 ggtgtcatct ttaagcttgc tctggtaggc gtcacatcc 16740
 accctccagg tcatcgtagg tgatcctaaa gcccctcgttgg 16800
 gccaacaag ttgcctaggc ggctgtgggtt gaaatccacc 16860
 gatgtctggc ttggaggcgtt cgtttgcata taccggc 16920
 gtcacccatcg agtctgaatgt tgccgggtgtc aaactttacc 16980
 cccgttctgc ctgcgcgtt tcaagttgtt ctcacgcgtt 17040
 ggtcatgttgc tggaggtagt tgccctcgttggc gacgcgc 17100
 ctccacatgtt ttgtccttagt taaggcggcgtt gacaccatc 17160
 ggttaacatgtt aactcgatcc tattttggcat gttggatgtc 17220
 gccccagtcg gaacgtcgta caagattgtt ggtctgtgtt 17280
 gtcattgttt tgaatgaccg tggtagaaa gttgtgtgtt 17340
 tgccacatcc gttgacttgtt tgcctccacaa gtcacacggc 17400
 caactcagag taacggatgc tttttctcc cccggtaggc 17460
 aaacggccggg tccaggggag catcgaaagg ggaacccaggc 17520
 gctcaccacg ctctcgtagg agggaggagg accttcctca 17580
 actaaggggaa atacaagaaa accaacgcgc ggtgccatgg 17640
 tgcacatgc tttttttttt tttttttaaa acattctcc 17700
 caaacgggtt gccactccct cccaaatcca ggacgtgtt 17760
 cctccacac cagaccccgca tgacgggtgtt gccttgcgtt 17820
 cgggcacatc cctgtgttcc tgccatcgtt tccacttaggc 17880
 tctctatccc gttgtggaa aatgcccggag gcagggttctt 17940
 gcgagttgtt tagtactcc tccatcgccca gcaggcgcggc 18000

taaaagaccc tatcaagctt gaaaaatgggc tactcgcatc tgaccgcggg gccgcagcgc 18060
 ctagatcgga caagctgctt ggcctgcgga agcttcctt tcgcagcgcgcc gcctctgcct 18120
 gctcgcgctg ttgcaactct agcagggtct gcgggttgcgg ggaaaacacg ctgtcgctta 18180
 tgtcgccccca gaggaatcca tcgttacacct cgggcacccaa aatcccccg gtgtagaaac 18240
 caggggggcggt tagccagtgc gggttcaaga tggcattggt gaaatactcg gggttcacgg 18300
 cggccgcggt atgcaagtagt tccattagc gattgataaa cggccgggtt gaggcataca 18360
 tgcccgggtt catgtgcgc gcggtcatgt ccagcgccac gctggcggtt accccgtcgc 18420
 gcatcaggtt aaggctcacg ctctgctgca catagcgcaa gatgcgcctcc tcctcgctgt 18480
 ttaaactgtg caacgagggg atcttctggc gccggttggt cagcaggtag ttcagggttg 18540
 cctccaggct gcccgtgtcc tcctgccccca gcgccgggtt gacacttgta atctcctgg 18600
 aagtatgctc gtccacatgc gcctgaccta tggcctgcgc gtacagtgtc agcaagtgac 18660
 ctaggtatgt gtcccgggac acgctgccccac tgcgtgtgaa gggcgctatt agcagcagca 18720
 acaggcgca gttggcggtc agcaagctag acacggcgcgc gcggtcgccct gtgggagccc 18780
 gcaccccccac cagccccctgc aagttcttga aagcttggct caggttacg gtctgcaggc 18840
 cttgtctact ggtctggaaa aaatagtcg gcccggactg gtacacccca ctttgcgggt 18900
 tctcagtcac cattagccgc agtgcgcctca caaatgggt gttagtcctcc tgcccccg 18960
 gcacgttggc gggctgtgtt ctcaggaagg cggttagtgc aaccatggag cccaggttgc 19020
 cctgtcgctg cgcgcgcctca cgctgcgcgc cggcctcgccg cacatcccc accagccgg 19080
 ccaggttgggt ctgcacgttgc cgcgttggt aacgagccac gcgctgaagc agcgcgtcgt 19140
 agaccaggcc ggcctcatacg gcgcggatgg ccctgttttgc ggccagcgcgc ttacagatcg 19200
 ccagcacctt ctgtgcgtg gggtttgcgc ggcgcggggac caccgcttcc agaattgcgg 19260
 agagccgggtt ggcctgcgcgc tgctgccccca acgcgtcagg gttacgcgc gtcagcgcaca 19320
 tgatgcggtc catgacacttgc cgcgcgttgc ccgtggagtt aaggccggac ggctggctct 19380
 gcagcgcgcgc cgcacccgcgc gggctcggtt cgtcttcgtatcatctgatca gaaacatcac 19440
 cgotttagtac tgcggctctt ctggctcgatccatcgatatt cctccacgcgc 19500
 gccgacgttgc cagcgcgcgc gggctgcacccgc cggcagccca ggtccggccc cagctgcctc 19560
 cagggcgcgtt cggctgggg cccagcgcgc gtcagcgcgc gctcaaaatg aggactcggc 19620
 ctctctatcg cgcgtccccgc tgccagccag ggccttttgc aggctgtgc tca gctcg 19680
 gtgcgtgagc tcgcggccgc ggctcacgtt cacggccttgc tggatgcgcgt cgttgcgata 19740
 aacgcccagg tcgtcgctca aggttaagcac cttaaacgcgc atgcgcatgt agaaccctctc 19800
 gatctttacc tccttgcgttca tgggaacgta aggggtatgg tataatcttgc gggcgtaaaa 19860
 ctggcccaga ctgagcatgg aatagttaat ggcggccacc ttgtcagcca ggctcaagct 19920
 gcgcctctgc accactatgc tctgcagaat gtttatcaaa tcgagcagcc accggccctc 19980
 gggctctact atgttttagca ggcgcacccctt gaatgcctcg ttgtccctgc tggtctgcac 20040
 tataaggaac agctgcgcacca tgagcggctt gctatttggg ttttgctcca ggcgcgttac 20100
 aaagtcccac agatgcacca gtcctatacg cacccctcg cgcgcaccaa gcgtgcgcac 20160
 gtgggtgtta aagctttttt gaaagttaat ctccctggttc accgtctgtc cgtacgcgg 20220
 taccaggtcg gcggccgcgc cgtgtgcgcgc cgcgggacta atcccggtcc ggcgcgtcgg 20280
 ctc当地tcc tcctcgccgc gcaaccgcgc gcggttccagg ccatgcccactcgcgc 20340
 tgcgtggaaac ttgcgtatccc gcatctccgc gggctcttgc ccctcgccgt cgcaaaacag 20400
 gttctggccgc ggcacgtacg ctcgcgcgt gtcacgcctt agtcgcaccc ttgggtgtcg 20460
 ctcaggagag ggctcgctca ggcgcgcgc gccctcgccccc tcctccaaatg ccaggttagtgc 20520
 cccggcccccgg cgcgcgggggg gttcgtaatcc accatctggc gccgcgtcgc ggcggatgt 20580
 tgcccccttcgt gacgggttag gagaaggggg ggggtgccttgcatgtctgc gctgtcttttgc 20640
 ctcttgcgcgc tgctgaggag gggggcgcatt ctggccgcgc accggatgca tctgggaaaa 20700
 gcaaaaaaagg ggctcgctcc ttgttccggat ggaatttgc aacgggggtct tgcatgacgg 20760
 ggaggcaaac ccccggtcg cgcagtcggc cggcccccggactcgtcaaccg ggggtcctgc 20820
 gactcaaccc ttggaaaata accctccggc tacaggagc gagccactta atgctttcg 20880
 tttccagcct aaccgcattac ggcgcgcgc gccagtgccgc aaaaaagcta gcgcagcgc 20940
 cgcgcgcgcct ggaaggaagc caaaaaggagc gctccccctgt tgcgtacgt cgacacactg 21000
 gggtcgacac gccggccggta accgcacatggatc acggccggatc cggccggatc 21060
 accccggctcg tccgcattca tacccttgcg aatttacca ccagaccacg gaagagtgc 21120
 cgcttacagg ctctcttttgcacggctca gagcgtcaac gactgcgcac gcctcaccgg 21180
 ccagagcgctc cgcaccatgg agcactttt gccgcgcgc aacatctggatc accgcgtccg 21240
 cgacttccgc cgcgcctccca ccaccgcgcgc cggccatcacc tggatgtccca ggtacatcta 21300
 cggatatcat cgccttatgt tggaagaccc gccccccggatc ccccccggccca cccctacgc 21360
 gcccctctac cgccagccgc cgccgcaccc tttgtggatc tatcagatacc tggtgcggac 21420
 ttgcaacgcac tacgtctttgc actcaaggccgc ttactcgctc ctcaggtaca ccgagctctc 21480

gcagccgggt caccagaccg ttaactggtc cgttatggcc aactgcactt acaccatcaa 21540
 cacgggcgca taccaccgct ttgtggacat ggatgacttc cagtctaccc tcacgcaggt 21600
 gcagcaggcc atattagccg agcgcgttgt cgccgacacgt gcccgccttc agccgatgag 21660
 gggcttcggg gtcacacgcg tgggaggaag agggcgccac ctacggccaa actccgcccgc 21720
 cgccttagcg atagatgcaa gagatgcagg acaagaggaa ggagaagaag aagtgcgggt 21780
 agaaaggcto atgcaagact actacaaaaga cctgcgcccga tgtcaaaacg aagcctgggg 21840
 catggccgac cgcctgcgca tttagcaggc cggacccaag gacatggtgc ttctgtcgac 21900
 catccggcgt ctcaagaccg cttactttaa ttacatcatc agcagcacct ccggccagaaa 21960
 caaccccgac cgccaccggc tgccgccccg cacggtgctc agcctacctt ggcactgtga 22020
 ctggttagac gcctttctcg agagggtttc cgatccgggtc gatgcggact cgctcagggtc 22080
 cctcggtggc ggagtaccta cacaacaatt gttgagatgc atcgtttagcg ccgtatccct 22140
 gccgcacggc agccccccgc caacccataa cggggacatg acggggggcg tcttccaact 22200
 gcccggccgc gagaacggcc ggcgcgtcac cgagaccatg cgccgtcgcc gccccggagat 22260
 gatcgagcgc tttgtcgacc gcctccgggt ggcgcgtcg ggcgcgtcg tccccccctcc 22320
 cccaccggcg ccagaagaag aagaagaagg ggaggccctt atggaagagg agattgaaga 22380
 agaagaggcc cctgttagcct ttgagcgcga ggtgcgcgac actgtcgccg agctcatccg 22440
 tcttctggag gaggagttaa ccgtgtcgcc ggcgaactcc cagttttca acttcgcccgt 22500
 ggacttctac gaggccatgg agcgccttga ggccttgggg gatatcaacg aatccacgtt 22560
 gcgacgctgg gttatgtact tcttcgtggc agaacacacc gccaccaccc tcaactacct 22620
 ctttcagcgc ctgcgaaaact acgcccgttt cgcggccgac gtggagctca atctcgcgca 22680
 ggtggtcatg cgccggccgc atgcccgaagg gggcgtggtc tacagccgcg tcttggAACGA 22740
 gggaggccgc aacgccttct cgcagctcat ggcgcgcaccc tccaacgacc tcgcgcac 22800
 cgtggagcga gccggacgcg gagatctca ggaggaagag atcgagcagt tcatggccga 22860
 aatcgccstat caagacaact caggagacgt gcaggagatt ttgcgcagg ccggccgtcaa 22920
 cgacaccgaa attgattctg tgcgactctc ttcaagggtt aagctcaccg gggccgtcg 22980
 cttcacgcg aggccggaga ttcaaggagat caacccggc gtcgtcgctg tcgcgcac 23040
 ccttcggcgcc cagcaccagg ttctggccgc ggcggccgc gacgtggccc tgccccctct 23100
 cccggccgggt cccgagccccc ccctacctcc gggggccgcg ccgcgtcacc gcttttagat 23160
 gcatcatcca aggacacccc cgcggcccac cgcggccgcg gcggtaccgt agtcgccgc 23220
 cggggatgcg gccttttgcg agtcatcgcg gcccgcacca accagccctt gaaaatcagg 23280
 tattcacctgg acctagcccg gcgcctgacc cggctatgcg aggtaaacctt gcaggagctc 23340
 cccctcgacc tgcggccgog ggagctccag accatggaca gctcccatct gcgcgatgtt 23400
 gtcatcaaggc tccgaccgcg ggcgcggac atctggactt tgggctcgcc cgccgtgggt 23460
 gtccgatcca ccataactcc cctcgagcag ccagacggtc aaggacaagc agccgaagta 23520
 gaagaccacc agccaaaccc gccaggcggag gggctcaaat tcccactctg ctcccttgcg 23580
 cgccgtcgctc aggtcaacct cgtcaggat gtacagcccg gcaactttttt gatcggtctcg 23640
 gcacgtttt acaaaaagcca gcacgagtgt tcggccgtc agttcttccc gatcggtctcg 23760
 cacatcaaca gccactccctc caactgggtt cgggagatcc agacctatac ttggatgggg 23820
 catccctcgca ccgagcgtct ctttgcacc tacatgttag gcctttggga agcagctcg 23880
 gccccttcgt gccccttcatttgcg ctgggttatga agttcgccgg agatgagcc 23880
 ctggtgaccg ccgcgcgaga cctagccgtg gaccctggat gggaccgcgt ggaacaagac 23940
 ccgccttaccc tctactgcattt caccggcggaa aaaatggcca taggtcgccca gtttaggacc 24000
 ttgcgcacc acctgcaaat gctaattggcc cgtgacctgt ggagctcatt cgtcgcttc 24060
 aacccttcattt ttgcagactg gcccctgtca gAACACGGGC tcagctcccc tgaggagctc 24120
 acctacgagg aactttaaaa attggcccttca atcaaggggca ccccgccgtt ctggaaactt 24180
 tacatcgtaggg gccacaacat caacggcttc gacgagatcg tgctcgccgc ccaggttaatt 24240
 aacaaccgtt ccgagggtgcc gggacccttc cgcatcacac gcaactttat gcctcgccgc 24300
 gggaaagatac ttttcaacga tgcgtccatc gcccgtccaa acccgccgtt caaaaagcgc 24360
 acggactttt tgctctggga gcaggccgaa tgcgacgaca ctgacttcaa ataccagtac 24420
 ctcaaaagtca tggtagggc cacctttgcg ctcacccaca cctcgctccg gaaggccgcg 24480
 caggcatacg cgctaccggc agaaaaaggaa tgctcgccct accaggccgt caaccagttc 24540
 tacatgctag gctcttaccg ttccggaggcc gacgggtttc cgatccaaga gtactggaaa 24600
 gaccgcgaag agtttgcctt caaccgcgag ctgtggaaaa aaaagggaca ggataagtat 24660
 gacatcatca aggaaaccct gactactgc gcccgtccaa tgcaaggctc cggccgagctg 24720
 gtcaacaaggc tgcgcgactc ctacgcctcc ttctgtcggt acgcggtagg tctcacagac 24780
 gccagcttca acgtcttcca gcgtccaaacc atatcatcca actcacatgc catcttcagg 24840
 cagatagtct tccgagcaga gcagccccggc cgttagcaacc tcggtcccgaa ccttcctcgct 24900
 ccctcgacgc aactatacga ttacgtcgcc gccagcatcc gcgggtggaaatgctaccct 24960

acatatcttg gaataactcag agagccccctc tacgtttacg acatttgcgg catgtacgcc
tccgcgtctca cccaccccat gccatggggt cccccactca acccatacga gcgcgccgtt 25020
gcccggccgcatggcagca ggcgttagac ttgcaaggat gcaagataga ctacttcgac 25080
gcccggccgtctgc tgcccggggt cttaccgtg gacgcagacc cccggacga 25140
gaccactac cgccattctg ttcgcgaag ggcggccgccc tctgtggac gacgcagcta 25200
ctacgcggag agtagccac cagcgttgac cttgtcaccc tgcacaaccg caacgagcgc 25260
gtgcacctgg tggccgacga ggcaccacc gtcttcccg aatggcggtg cggggccg 25320
gaatacgtgc agctaaacat cgccccaag gagcgcggcc atcgcgacaa aacaggccg 25380
ctgcgtctca tcgccaagt gctgtccaaac gcccctctacg ggtcggttgc aaaccaaacc 25440
gacaacaaaa agattgtctt ttctgaccag atggacgcgg ccacccctcaa 25500
gcggggccagg tgaatatacaa atccctctcg ttttggaaa ctgacaatct 25560
gtcatggccg ctttgagag ggagtactca ccccaacagc tggccctcgc tagcgccgaa 25620
gcggaaagaga gtgaggacga acgcgcccccc ccccttttta atagcccccc 25680
cccggtcacg tgccctacac ctataaacca atcaccttcc ttgatggcga ttcaggaaaca 25740
atgtgtcttc acaccctgga gcgagtggac cccctagtgg acaacgaccg 25800
cacttagcct cttcgtgt ggcctggacg cgagccttcg tctcagagtg ctaccatcc 25860
ctatacggagg aggaccgcgg aacaccgctc gaggacaggc ctctcaagtc tgtatacggg 25920
gacacggaca gcctttcgt caccgagcgt ggacaccggc tcatggaaac 25980
aaacgcatca aaaagcatgg gggaaacctg gttttgacc ccgaacggcc 26040
tggctcggtt aatgcgagac cgtctcgccc gcctcgccg cggatgccta 26100
tcggtatttc tcgcgccccaa gctctacgccc cttaaaagtc tgcactgccc 26160
gcctcctcca agggcaagct ggcgcacaag gcccacggc cggaggggct 26220
accatggtca aatgtctacct ggccgacgcg cagggcgaag accggcagcgc 26280
agcaggacca gcctcaagcg caccctggcc agcgcgcagc cggagcgc 26340
gtgaccggaga ctacgctgac gaggaccctg cggccgtgga aagacatgac 26400
ctggacgagc accgactact gccgtacagc gaaagccgccc 26460
atatgctgga tcgagatgccc gtagagcacg tgaccgagct gtgggaccgc 26520
ttggtcaaaac gctcaaaagg atgcctacgg cggacggccct 26580
cttccttgca agaactgcta tcgctggcgc caaacctgtt 26640
aaaacatgca agtcaggggac atgctaaccg aagtggcccc cctgctcagg 26700
gctgcagctc tcttaactac cagttgcagc cggtaatagg tlgatattac 26760
gctcggttaa gtcgcagctg ctcaggaacc tgctttttc ccagctgtatc 26820
cgaaaacgggt ttcttcatc gcccccgagg tagacatgat ccccccatct 26880
cgtggaaaat gcaaattctgt gaggtaact acgccccctgg gccggatgga 26940
cgcagcttgg cacccctccgc cccgcgtttg taaaaatggc ctatgacgat 27000
aacacaacta tgacgttagt gatcccagaa atatcttcgc ccaggccgccc 27060
ccattgccccat cattatggac gaaatgcatgg aaaatctcg 27120
agttcttcca cgcatttcct tctaagctac atgacaaatt tcccaagtgc 27180
ctgtgcttgggt ggttctgcac aacatgaatc cccggagggg 27240
acctaaaaat acagtccaaag atgcattctca tatccccacg 27300
accgctttgtt aaacacttac accaaggggcc tgcccttggc 27360
acatttttag gcaccacgcgc cagcgtctct gctacgactg 27420
cgcagcatga agctctgcag tggtgctacc tccacccctag 27480
atctgaacat ccagagtccac ctttaccacg tcctggaaaa aatacacagg 27540
accgagaccg ctggtcccggt gcctaccgcg cgcgcääaaac 27600
acacttgctt gatcaaaatc caaacagagt ctggttttt 27660
gggaggggag gaagcttca gggcagaaac ctgctggcgc agatccaaca 27720
aacgcacatta atttcccggt tcaaaagaatc caattgtgccc 27780
atcgcggccg gatgaacggg aagctgcact gcttgc 27840
agtcacaatc cccggggccg tggctgcagc ggctgaagcg 27900
caacggcggtt ccagacacgg tctcgtaggt caaggtagta 27960
gacgcacatca atgctggagg ccatcacatt ctgacgcacc 28020
cggtgtctaa tatgagctca caatgttcc atcaaacgag 28080
tgctgcaaaa cagatacaaa actacatgag acccccacct 28140
taagccccgc ccatcgatgg caaacagcta ttatgggtat 28200
ggttgccccg tattcagtgt cgctgatttg tattgtctga 28260
gatgcagatc attaataacg atacctgcgt cataattgt 28320
cctccacgcga cgttgtgata ttagatgat aatcattatc 28380
acttacggg 28440

tgatccgaca ggtaacgggg cggcgacctc gcgggtttt gctatttatg aaaattttcc 28500
 gtttaaggc gtttcgttc ttcttcgtca taacttaatg ttttattta aaataccctc 28560
 taaaaagaaa ggaaacgaca ggtgctgaaa gcgaggctt ttggcctctg tcgtttcctt 28620
 tctctgttt tgcgttgaa atgaacaatg gaagttaacg gatccaggcc gcgagcaaaa 28680
 gcccagcaaa aggccaggaa ccgtaaaaag gccgcgttgc tggcgtttt ccataggctc 28740
 cccccccctg acgagcatca caaaaatcaa cgctcaagt agaggtggcg aaaccgcaca 28800
 ggactataaa gataccaggc gtttccccc ggaagctccc tcgtgcgtc tcctgttccg 28860
 accctgccgc ttaccggata cctgtccggc tttctccctt cgggaagcgt ggcgccttt 28920
 catagctcac gctgttaggt ttcagttcg gtgttaggtc ttgcgtccaa gctgggctgt 28980
 gtgcacgaac cccccgttca gcccgcacgc tgccgccttat ccggtaacta tcgtctttag 29040
 tccaaccggg taagacacga cttatcgcca ctggcagcag ccactggtaa caggattagc 29100
 agagcgaggt atgttaggcg tgctacagag ttcttgaagt ggtggcctaa ctacggctac 29160
 actagaagaa cagtattgg tatctgcgt ctggccaaagc cagttacctt cggaaaaaaga 29220
 gttggtagct cttgatccgg caaacaaacc accgctggta gcggtggtt ttttgggttc 29280
 aagcagcaga ttacgcgcag aaaaaaagga tctcaagaag atccttgat cttttctacg 29340
 gggctcgacg ctcaagtggaa cggaaactca cgttaaggga ttttggtcat cagattatca 29400
 aaaaggatct tcacctagat ccttttaaat taaaaatgaa gttttaaattc aatctaaagt 29460
 atatatgagt aaacctggtc tgacagttac caatgctta tcagtgaggg acctatctca 29520
 gcgatctgtc tatttcgttc atccatagtt gcctgactcc ccgtagtgt aataactacg 29580
 atacgggagg gcttaccatc cggcccccagt gctcaatga taccgcgtga cccacgctca 29640
 cccgctcctg atttatcagc aataaaaccag ccagccggaa gtgccgagcg cagaagtgt 29700
 cctgcaactt tatccgcctc catccagtct attagttgtt gccgggaagc tagagtaagt 29760
 agttcgccag ttaatagttt tcgcaacggtt gttgccattt ctacaggcat cgtgggtgtca 29820
 cgctcgctcg ttggatggc ttcattcagc tccgggtccc aacgatcaag gcgagttaca 29880
 tgatccccca tggatggcaaa aaaaagcgtt agtccttcg gtccctccgt agttgtcaga 29940
 agtaagttgg ccgcgtgtt atcaactcatg gttatggcag cactgcataa ttctcttact 30000
 gtcatggccat ccgttaagatg cttttctgtt actgggtgagt attcaaccaa gaataacggga 30060
 taataccgcg ccacatagca gaactttaaa agtgcctatc attgggaaac gttcttcggg 30120
 gcgaaaaactc tcaaggatct taccgcgttt gagatccagt tcgatgtaa ccactcgcc 30180
 acccaagtga tcttctgtat cttttacttt caccagcggtt tctgggttag caaaaacagg 30240
 aaggcaaaat gccgcaaaaaa agggaaataag ggcgacacgg aaatgttcaa tactcataact 30300
 ttccctttt caatattttt gaagcatatc tcagggttat tgtctcatca gcggatacat 30360
 atttg 30365

<210> 5
 <211> 33
 <212> DNA
 <213> Homo sapiens

<400> 5
 gcggaattcg gcttggtgac ttagagaaca gag 33

<210> 6
 <211> 33
 <212> DNA
 <213> Homó sapiens

<400> 6
 gcgggatcct tgaacccgga ccctctcaca cta 33

<210> 7
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>

<223> derived from Adenovirus

<400> 7
actctttcc gcatcgctgt 20

<210> 8
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> derived from Adenovirus

<400> 8
cttgcgactg tgactggta g 21

<210> 9
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> derived from Adenovirus

<400> 9
ccgcacccac tatcttcata 20

<210> 10
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> derived from Adenovirus

<400> 10
ggtgtccaaa ggttcggaga 20

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US01/27682

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :C12N 15/86, 15/861, 5/10, 15/11, 15/68, 15/64, 15/65; A61K 48/00
US CL :Please See Extra Sheet.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 424/93.1, 93.2, 93.6; 435/320.1, 69.1, 455, 456, 457, 325, 369, 91.1, 91.4, 91.42

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No |
|-----------|--|----------------------|
| X | US 6,080,576 A (ZAMBROWICZ et al.) 27 June 2000 (27.06.00), see entire document, especially Figure 2, claims 1-3 and columns 10 and 16. | 1-2, 4-11, 13 |
| A | US 5,919,676 A (GRAHAM et al.) 06 July 1999 (06.07.99), see entire document, particularly claims 1-6 and columns 2 and 5-6. | 1-34 |
| A | AGAH et al. Gene recombination in postmitotic cells. Journal of Clinical Investigation. July 1997, Vol. 100, No. 1, pages 169-179, especially pages 171-173. | 1-34 |

Further documents are listed in the continuation of Box C. See patent family annex.

| | | | |
|-----|---|-----|--|
| * | Special categories of cited documents: | "T" | later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention |
| "A" | document defining the general state of the art which is not considered to be of particular relevance | | |
| "E" | earlier document published on or after the international filing date | "X" | document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone |
| "L" | document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) | "Y" | document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art |
| "O" | document referring to an oral disclosure, use, exhibition or other means | "R" | document member of the same patent family |
| "P" | document published prior to the international filing date but later than the priority date claimed | | |

Date of the actual completion of the international search

01 NOVEMBER 2001

Date of mailing of the international search report

31 DEC 2001

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703) 305-8280

Authorized officer

David Guzo

Telephone No. (703) 305-0196

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US01/27682

A. CLASSIFICATION OF SUBJECT MATTER:
US CL :

424/93.1, 93.2, 93.6; 435/320.1, 69.1, 455, 456, 457, 325, 369, 91.1, 91.4, 91.42

B. FIELDS SEARCHED

Electronic data bases consulted (Name of data base and where practicable terms used):

WEST, Dialog, NTIS, Medline, Biotech, Biosis, Biosci, Chemical Abstracts
Search terms: adenovirus, recombinase, target site, Cre, cleavage site, gene therapy